

# 관찰의 이론 적재성을 전제했을 때의 Ad Hoc 가설 개념 연구\*

이찬우  
(서울대 철학과)

## 1. 들어가며

일반적으로 “ad hoc 가설”이란 특정한 이론에 부합하지 않는 관찰이 이루어질 경우, 오로지 그 관찰이 이론에 부합하게끔 하는 목적을 위해서만(“ad hoc”) 도입되는 보조가설*auxiliary hypothesis*의 한 유형을 가리킨다. 예컨대 가설-연역 모형*hypothetico-deductive model*을 본으로 취하여, 이론 T와 관찰 E에 관하여 ‘ $T \rightarrow E$ ’이 타당하지 않다고 전제하자; 이때 보조가설 A가 오로지 ‘ $((T \& A) \rightarrow E)$ ’를 타당하게끔 하기 위해서만 도입되는 경우, A를 흔히 “ad hoc 가설”<sup>1)</sup>이라고 부른다.

ad hoc 가설을 도입하여 이론을 방어하는 것은 대개 부정적인 것으로 간주된다: 예컨대 ‘차고 안의 용은 사람 눈에 보이지 않는다’, ‘차고 안의 용은 체온이 주변 환경과 일치해서 적외선을 통한 탐지가 불가능하다’와 같은 보조 가설들을 연거푸 들어 ‘우리 집 차고에 용이 산다’는 이론을 방어하는 것은 ad hoc 가설이 부정적으로 쓰이는 대표적인 예화다.<sup>2)</sup>

---

\* 이 글은 필자가 2014년도 1학기에 제출한 학사학위 논문의 표현 및 인용 방식 등을 일부 수정한 것이다.

1) “ad hoc hypothesis”는 “에드혹 가설”, “임시방편 가설”, “임시변통 가설”, “미봉 가설”, “편의적 가설” 등으로 다양하게 번역되며, 아직 확립된 한국어 번역어는 없는 것 같다. 따라서 본고에서는 잠정적으로 “ad hoc 가설”이라는 표현을 사용하겠다.

2) Carl Sagan, *The Demon-Haunted World: Science as a Candle in the Dark*,

반면 천왕성 궤도가 뉴턴 역학에 합치하지 않는다는 관찰을 설명하기 위하여 해왕성의 존재를 가정한 것은 종종 보조 가설의 대표적인 성공 사례로 언급된다. 해왕성은 천왕성 궤도에 섭동을 일으키는 가설적 천체로 도입되었지만, 나중에 실제 관찰을 통하여 입증되었기 때문이다.<sup>3)</sup>

ad hoc 가설은 이론 시험/선택의 과정에서 장애를 야기한다. 상기 예시들에서 나타나는 바와 같이, 이론과 상치되는 관찰이 이루어질 때에도 보조 가설을 도입함으로써 반증*falsification*을 차단할 수 있기 때문이다. 따라서 이는 기존의 잘못된 이론이 반박되는 것이 아니라, 그때 그때 이상현상이 발생할 때마다 대중적인 보조가설의 도입만으로 유지되는 문제를 낳을 수 있다. 이러한 문제는 “어떠한 진술도 체계 내에 큰 수정만 가해진다면 참으로 여겨질 수 있다”(Quine, 1951)는 이른바 뒤엠-콰인 논제*Duhem-Quine thesis*의 극단적 예시에 해당한다.

하지만 ad hoc 가설을 “관찰이 이론에 부합하게끔 하는 목적을 위해서만 도입”하는 것으로서 정의하는 동기주의적 해석은 전적으로 과학자의 의도에 의존하므로 이론 시험/선택 과정에서의 준거로서 신뢰성을 확보하기 힘들다(Scott, 2012). 이를 막기 위해선 임의의 보조가설이 ad hoc 가설에 해당하는지 여부를 판단할 수 있는 조작적 판단 기준이 필요하다. 뒤에서 자세히 살피겠지만 ad hoc 가설을 판단하는 전통적인 기준으로는 크게 관찰 가능성과 이론 단순성이 거론되며, 이러한 기준들의 채택 여부는 관찰의 이론 적재성을 받아들이는지 여부에 따라 좌우된다.

---

Ballantine, 1996

- 3) 이때 해왕성 가설 같은 ‘긍정적’ 보조가설 또한 ad hoc 가설에 포함하여 분류하는 경우도 있다. 예컨대 Strevens (2001)는 상기한 모든 예시들을 ad hoc 가설의 범주에 포함시키며, 그 안에서 긍정적인 사례를 “glorious rescue”, 부정적인 사례를 “desperate rescue”라고 따로 분류한다. 본고에서는 ad hoc 가설의 부정적인 경우를 중점적으로 다루므로, “ad hoc 가설”은 기본적으로 부정적인 ad hoc 가설만을 지칭한다.

**관찰의 이론적재성** *theory-ladenness of observation* (이하 “이론적재성”) 논제란 과학자의 관찰 내용이 그 과학자가 견지하는 과학 이론에 의해 영향을 받는다는 주장이다; 이는 곧 관찰 자료가 이론들에 대하여 반드시 중립적인 것은 아니라는 점을 함축한다. 이론적재성의 수용은 과학의 객관성에 심각한 위협을 가져오는 것으로 받아들여지고는 하므로, 그 수용이 합당할지 여부는 과학철학에서 주요한 학적 쟁점에 해당한다(Franklin et al., 1989; Schindler, 2013a).

ad hoc 가설에 관한 기존의 많은 견해들은 관찰의 객관성을 전제하며 이를 ad hoc 가설 판단 및 이론 시험/선택의 기준으로 활용한다. 하지만 이론적재성을 받아들이면 ad hoc 가설에 관한 기존의 논의들은 더이상 유효하지 않다. 그래서 이론적재성을 받아들이는 경우, ad-hoc 가설 개념은 ‘단순성’이라는 형식화될 수 없는 심미적 가치에 의존하는 것으로 흔히 치부되고는 한다.

본고에서는 이론적재성을 도입하는 경우조차도 'ad hoc 가설'에 대해 여전히 이론-내재적인 개념화가 이루어질 수 있음을 보이고자 한다. 이는 '단순성'에 관한 대안적 형식화 방안을 제시하는 것을 통해 이루어진다. 본고에서는 먼저 고전적인 과학철학 문헌들의 검토를 통해 이론적재성과 ad hoc 가설 개념간의 관계를 보이고(2절), ad hoc 가설에 관한 현대의 여러 논의들을 살피며 이론적재성을 비롯한 여러 조건들의 부합 여부를 보이고자 한다(3절). 최종적으로 본고에서는 ad hoc 가설 개념에 대한 대안적 정의로서 ‘이론의 시간 복잡도’ 개념을 제시하고 이에 대한 정당화를 펼치도록 하겠다(4절).

## 2. 고전적 논의에서의 ad-hoc 가설과 이론적재성 개념

상기한 바처럼 ad hoc 가설 개념은 과학철학의 이론 시험/선택 과정에서 문제를 초래할 수 있는 잠재적인 요소이므로 과학철학의 여러 고전적 이론에서 흔히 다루어지는 주제다. 본고에서는 먼저 이러한 고전적 이론들에서 ad hoc 가설 개념이 다루어지는 방식들을 조명해 봄으

로써 그 도입 동기 및 접근 방식들을 확인해보고자 한다.

칼 험펠은 ad hoc 가설을 분간하는 명확한 기준은 없되, 이는 그저 과학사의 사례들에 비추어 판단될 수 밖에 없는 것이라고 주장한다 (Hempel, 1966). 다만 험펠은 ad hoc 가설들이 보이는 유의한 특징으로 ‘새로운 현상을 예측하지 못한다’는 점과 ‘계속된 보조가설의 도입으로 인해 전체 이론 체계가 지나치게 복잡해진다’는 점을 밝히며, 이 두 가지 특징을 지침*guidance*으로 삼아 ad hoc 가설 여부를 판단할 것을 제안한다(ch 3.4).

이로부터 ad hoc 가설의 개념화에 대한 두 가지 기조를 확인할 수 있다:

1. **관찰 가능성**: 보조가설 A가 이론 T에 부가될 때, A가 새로운 유관한 경험적 현상을 예측하지 못할 경우, A는 ad hoc 가설이다.
2. **이론 복잡도(단순성)**: 보조가설 A가 이론 T에 부가될 때, A와 그 이후에 도입되는 보조가설들로 인해 T가 지나치게 복잡해질 경우, A는 ad hoc 가설이다.

칼 포퍼, 임레 라카토슈 등 ‘반증주의자’로 분류되는 고전적 이론가들은 ad hoc 가설에 대하여 기조 1에 가까운 견해를 지니는 것으로 보인다.

포퍼는 ad hoc 가설의 도입이 가능하다면 일체의 반증적 경험이 아무런 논리적 하자 없이 거부될 수 있으므로, ad hoc 가설을 반증주의에 대한 심각한 위협으로 인식한다(Popper, 2002[1959], 19-20). 하지만 그는 규약주의자*conventionalist*들이 이론 시험/선택의 기준으로 ‘단순성’이라는 특성을 거론하는것에 반대한다; 규약주의에서 말하는 단순성이란 심미적/실용적인 가치로, 엄밀하게 사용되지 않은 탈-논리적 *extra-logical*인 가치라고 보기 때문이다(60,122). 포퍼는 보조가설의 좋고 나쁨은 해당 보조가설이 단독으로 시험가능한지에 달렸다고 설명하며(62), ad hoc 의 정도*degrees of ad hocness*는 시험가능성/유의성과 역의 관계에 놓인다고 명시한다(Popper, 1959) .

그러나 포퍼가 ‘미래에 발견될 것으로 추측되는 이론 체계의 논리적 귀결’에 호소하는 방식으로 보조가설을 도입하는 것에 반대하는 것과 달리(Popper, 2002[1959], 62), ‘세련된 반증주의자’를 자처하는 라카토슈는 보조가설이 현 시점에 시험가능성을 결여한다고 해서 즉각적으로 포기해서는 안된다고 주장한다(Lakatos, 1980). 라카토슈는 ‘ad hoc 가설’을 그 가설의 도입을 통해 부가되는 경험적 내용이 부재하는, 혹은 그 내용이 경험적으로 용인될 수 없는 가설로 정의한다(88). 하지만 그는 막스 플랑크의 복사 법칙 또한 당대엔 전적으로 임의적이고 형식적인 이론적 장치에 불과했음을 지적하며(80), 경험 내용 유무의 판단 시기는 유예할 수 있다고 주장한다. 이론의 반증은 즉시적 합리성 *immediate rationality*에 따르는 것이 아니라, 오로지 더 나은 반대 이론이 출현할 때에만 이루어지기 때문이다(35).

이론 시험/선택 과정에 있어서의 견해차에도 불구하고 포퍼와 라카토슈는 해당 가설과 결부되는 경험적 내용의 유무/정도차를 공통적으로 ad hoc 가설의 기준으로 제시한다. 이러한 입장들은 상기 분류에서 ‘관찰가능성’을 부각시키는 기조 1에 부합한다고 볼 수 있다.

하지만 이론적재성을 전제할 경우, 기조 1에 해당하는 견해들은 유지되기 힘들다. **관찰의 이론적재성**이란 곧 개별 이론에 영향을 받지 않는 중립적 관찰이 있을 수 없다는 것이기 때문이다:

이론적재성은 관찰 용어 *observational term*들이 이론적 전제에 영향을 받는다는 의미론적 이론적재성 *semantic theory-ladenness*과 아예 인지적 층위에서 감각 내용이 영향을 받는다는 감각적/심리적 이론적재성 *perceptual/psychological theory-ladenness* 등으로 구분될 수 있다(Elgin, 2008; Schindler, 2013b). 의미론적 이론적재성을 받아들일 경우, 상이한 이론을 지니는 두 관찰자는 중립적 관찰 용어의 부재로 인하여 동일 현상으로부터 상이한 관찰 문장을 얻는다. 따라서 두 이론은 중립적인 관찰 문장에 의한 이론의 반증 여부 판단을 기대할 수 없다. 나아가 감각적 이론적재성을 받아 들일 경우, 두 관찰자는 아예 상이한 경험 내용을 가지게 된다. 의미론적 이론적재성을 받아들일 땐 그나마

관찰자 언어들간의 번역의 가능성을 타진해볼 수 있는 것과 달리, 감각적 이론적재성에 따르면 경험 내용은 객관적 신뢰성을 가질 수 없으므로 이론 시험/선택 시에 준거로서 전적으로 신뢰될 수는 없다.<sup>4)</sup>

포퍼가 관찰의 이론 적재성을 인정했다는 점에서도 나타나듯(Thornton, 2013), 관찰가능성을 기준으로 제시하는 것이 원칙적으로 이론적재성과 배치되는 것은 아니다. 그러나 이러한 견해는 다음과 같은 비판에 직면한다:

- 이론이 “진실로 시험가능하다면 *genuinely testable*” 관찰을 기술하는 명제의 진위를 따질 수 있어야한다; 그렇지 못하다면 해당 명제는 반증 사례로서 정당화될 수 없다. 따라서 이론적재성을 전제할 경우, 반증 사례를 정당화시키는 것은 불가능한 듯 하다. 포퍼는 이러한 정당화가 “자유로운 결정”에 의해 이루어진다는 흡사 규약주의적인 대답을 내놓는다. 하지만 이는 그 자신의 본래 입장과 충돌할 뿐 아니라, “진리를 향한 진보적 접근”이라는 포퍼의 과학관에 어긋난다(Thornton, 2013, 9.1)
- 프톨레마이오스 천문학이 코페르니쿠스 천문학과 의 경합 과정에서 주전원 가설의 거듭된 도입을 통하여 스스로 이론을 교정하고자 했던 것은 과학사상 ad hoc 가설이 쓰인 것으로 여겨지는 전형적 예시이다(Scott, 2012). 하지만 주전원 가설은 일정한 궤도를 기술한다는 점에서 명백하게 시험적 귀결을 지닌다. 따라서 이는 포퍼의 견해에 대한 반례이다.
- ‘원리적으로 관찰불가능’하다는 것은 사실상 없다는 견해가 제기된다: 예컨대 크리스토퍼 헛트는 코페르니쿠스 천문학의 연주시차 문제를 들며 ‘원리적 관찰가능성’의 기준 자체가 시대에 따라 근

4) Elgin(2008)은 이론적재성이 타당하다 한들 이는 과학자 집단이 이론간의 공통된 척도를 찾지 못한다는 것을 함축할 뿐, 객관적인 “신의 눈”에서 볼 때는 여전히 이론간에는 공통된 척도가 있다고 주장한다. 하지만 과학 발전 과정에서 공약불가능성이 거론되는 것은 과학자(집단)의 한계를 설명하기 위한 인식론적인 동기에 기인한다. 따라서 Elgin이 (인식이 불가능한) 이론간 공통 기준의 실재를 주장하는 것은 본 공약불가능성 논증에 영향을 주지 못한다.

본적으로 달라진다고 주장한다(Hunt, 2012).

포퍼의 견해는 위와 같은 문제들에 대해 결정적인 답변을 주지 못하는 것으로 보인다. 또한 라카토슈의 경우에도 그저 문제를 유예할 수 있을 뿐, 경험 내용에 의존하여 두 이론간의 우열을 파악하는 것은 불가능하다는 문제를 여전히 노출시킨다.

반면 이론적재성의 요소를 보다 적극적으로 수용하는 이론가들로는 대표적으로 파울 파이어아벤트와 토머스 쿤을 들 수 있다.

파이어아벤트는 경험적 사실들이 중립성을 띠고 있으며 어떤 이론을 지니건 접근가능하다는 요지의 자율성 원리(*autonomy principle*)가 불완전함을 지적한다(Feyerabend, 1993, 26-7). 과학자들은 감각기관이 기만당할지도 모른다는 점을 감안해 경험 내용을 있는 그대로 신뢰하지 않는다(22), 더욱이 경험적 내용은 관찰용어, 보조가설, 배경지식 등의 요소들에 의하여 오염되어있기 때문이다(52).

그리고 쿤은 이론적재성을 관찰자들 각자가 놓인 세계(world)간의 차이로 설명한다: 서로 다른 패러다임에 속하는 두 관찰자는 동일 위치/시점에서도 서로 다른 대상을 본다(Kuhn, 1970, 150). 왜냐면 패러다임 내에서 이루어지는 과학 연구(“정상과학”)란 곧 관찰대상인 자연을 미리 마련한 개념적 틀에 적재하려는 활동이기 때문이다(5). 쿤은 관찰자들의 게슈탈트(*gestalt*)가 비자발적으로 교체된다는 심리학적 연구에 근거하여 관찰 내용이 관찰자의 패러다임에 따라 달라질 수 있다고 주장한다(85, 111-120).<sup>5)</sup>

두 저자 모두 과학적 발전이 곧 진리에 접근해 가는 과정인 것이 아

5) Hoyningen-Huene(1993)는 Kuhn(1970)에서 “세계”가 맥락에 따라 “세계-그-자체”로도, 혹은 “현상 세계”로도 새겨질 수 있는 중의어임을 지적하며, 이러한 용례들은 쿤 후기 철학에 들어 ‘자극(*stimulus*)과 감각(*senstation*)’이라는 개념어들로 각각 분리된다고 주장한다(ch 2.1, 2.2). Hoyningen-Huene은 이러한 견지와 칸트 초월 철학 간의 유사성을 지적하기도 한다(ch 2.1). 흥미롭게도 칸트와의 유사성은 Feyerabend(1962)에서도 자인된다.

님을 강조한다는 점은 이들이 이론적재성의 요소를 수용한다는 것을 단적으로 예증한다(Feyerabend, 1993, 21; Kuhn, 1970, 206; cf. Hoyningen-Huene, 1993, ch.7.6.c). 아울러 위 논저들은 두 저자가 이론간의 공약불가능성 *incommensurability*을 주장하는 근거가 되기도 한다; 쿤은 상기한 세계간의 차이를 공약불가능성의 세 가지 이유 중 가장 근본적인 것으로 지목하였고(Kuhn, 1970, 150), 파이어아벤트는 관찰용어들의 이론적재성으로 인하여 이론간을 매개할만한 중립적 기반이 부재할 수밖에 없다고 주장하였다(Feyerabend, 1962; Oberheim and Hoyningen-Huene, 2013).<sup>6)</sup>

하지만 파이어아벤트와 쿤은 이론간의 공약불가능성이 곧 이론간의 비교불가능성을 함축하지는 않는다는 점, 즉 이론 시험/선택 과정이 전적으로 비합리적인 것은 아님을 역설한다(Oberheim and Hoyningen-Huene, 2013). 예컨대 정상과학에서 이상현상 *anomaly*이 출현함에 따라 패러다임에 위기 *crisis*가 점차 도래하고(Kuhn, 1970, 68), 문제 해결 역량 등을 두고 이루어지는 경쟁 패러다임의 경합 끝에 과학 혁명이 성취된다는 쿤의 모형은 “체계적 결정 절차”(200)는 아닐지라도 과학 발전에 대한 합리적 설명을 제공한다.

다만 뒤엠-콰인 명제의 문제제기에서도 나타나듯, 쿤의 저작(Kuhn, 1970)에서 문제가 되는 지점은 패러다임 경합 과정 가운데 양 패러다임 모두 관찰자료에 대한 순환성을 지니며, 실험적/논리적 수단만으로 우열을 판가름할 수 없다는 점이다(94). 이를 가능하게 하는 요인으로 쿤은 기존 패러다임과 이상현상을 합치시키기 위하여 이론에 대해 *ad hoc*한 수정이 가해지는 점을 지목한다(78, 83). 프톨레마이오스 천문학이 코페르니쿠스 천문학과의 경합 과정에서 주전원 가설의 거듭된 도입을 통하여 스스로 이론을 교정하고자 했던 것은 그 대표적인 예시다(68, 181)<sup>7)</sup>.

6) Hoyningen-Huene(1993), Brown(2005) 등은 쿤의 이론이 후기에 접어들면서부터는 감각적 이론적재성에 관한 옹호를 철회하고, 경험적 자료의 객관성 등을 보다 강하게 주장하는 방향으로 나아간다고 주장한다. 하지만 여기서는 이른바 ‘쿤의 전기 이론’에 해당하는 Kuhn(1970)의 서술에 초점을 기울이겠다.



이러한 문제를 타개하기 위하여 쿤은 이론에 ad hoc한 수정이 가해짐에 따라 이론 복잡도가 급격하게 증가하는 점을 주목한다. 쿤은 중근세 천문학자들이 프톨레마이오스 체계의 복잡성을 두고 이론의 진위를 의심한 사례들을 인용하며, 정상과학 이론이 지나치게 복잡함에 따라 초래되는 문제 풀이 과정의 파탄이야말로 정상과학 위기의 핵심이라고 주장한다(69)<sup>8)</sup>.

쿤의 상기 견해는 이론적재성을 받아들인 입장에서 ad hoc 가설에 대한 개념화가 어떻게 이루어져야하는지를 잘 보여주는 예시이다: ad hoc 가설의 도입은 위기에 봉착한 패러다임이 이상현상을 무마함으로써 이상과학*extraordinary science*과의 경합 가운데 정상과학의 지위를 유지하고자 하는 반동적 시도다. 그러나 ad hoc 가설의 도입은 이론의 복잡도 증가를 초래하므로, 연구자가 이를 인식함으로써 ad hoc 가설에 대한 판별과 배제가 이루어진다는 것이 쿤의 주장이다. 이러한 입장은 본 절의 서두에서 제시한 ad hoc 가설 개념화의 두가지 기조 가운데 기조 2에 부합한다.

- 
- 7) “이론적재성을 전제한다면 애초에 이론에서 벗어나는 이상현상이 발견될 수 없는 것 아닌가?”라는 문제 제기가 가능하다. 이에 대하여 Hoyningen-Huene(1993)은 세계-그-자체가 관찰자-종속적인 현상세계에 저항함으로써*resist* 관찰 자료에 구속력을 행사한다는 해석을 제시한다(268). 이는 관찰의 이론적재성을 용인하는 대개의 견해들이 ‘이론에 의하여 관찰이 전적으로 결정된다’는 것보다는 더 약한 입장을 견지한다는 점과 맥락을 같이한다(Schindler, 2013a). 그럼에도 불구하고 관찰에 대한 이론의 영향을 배제할 수 없는 이상, 서로 다른 이론을 받아들이는 관찰자들의 경험 내용이 동일하다는 것은 전혀 보장되지 않는다.
- 8) Feyerabend(1993)는 지동설(77, 104-105), 뉴턴(44-45) 및 현대의 수리물리학(49) 등을 ad hoc 가설의 대표적인 사용례로 들며, 이들이 “(주어진) ‘사실들’과 추후 보완을 통해 이들을 설명할 수 있을 것으로 예상되는 견해들이 맞닿을 수 있는 잠재적 공간을 마련해준다”고 설명한다(157). 그러나 파이어아벤트가 쓰는 “ad hoc 가설”의 용례는 각주2에서 언급한 광의의 “ad hoc 가설”에 해당하는 것으로 보인다; 이는 파이어아벤트가 라카토슈의 ad hoc 가설에 관한 견해를 소개하길, “[퇴행적이지 않는한, ad hoc 가설의 도입을] 금지하지 않는다”라고 말하는 것과 합치한다(158, 각주11). 따라서 파이어아벤트의 “ad hoc 가설” 용례는 본고에서 규정하는바 보다 넓다.

이러한 기준은 쿤에 의해서도 명시적으로 언급된다: 쿤은 이상과학 패러다임이 정상과학 패러다임을 대체하게 되는 요건 중 하나로 이전 패러다임보다 더 간결*neat*/적합*suitable*/간단*simple*해야 한다는 점을 언급하며(155), 또한 “패러다임” 개념이 보다 정밀하게 정의되는 1969년의 후서*postscript*에서 이론의 단순성은 패러다임을 구성하는 가치중 하나로 거명된다.

쿤은 이론적 단순성/복잡도란 그 중요성에도 불구하고 미학적인 개념일 따름이며, 주관적/개인적/불명확할 수밖에 없는 척도라고 간주한다(155-158). 하지만 ad hoc 가설 도입이 이론 시험/선택에 미칠 수 있는 부정적 영향을 고려할 때, 그 척도에 대하여 이처럼 상대주의적 견해를 보이는 것은 쿤이 자기 이론이 합리적이라고 강변하는 것에 잘 부합하지 않는 것처럼 보인다.

쿤이 봉착하는 결론과 달리, 본고에서는 관찰의 이론적재성을 전제할 때에도 ad hoc 여부를 판단할 수 있는 대안적 개념화 방식을 찾아보고자 한다. 이를 위하여 3절에서는 ad hoc 가설에 관한 현대의 여러 연구들을 살펴보고, 이론적재성과의 합치 여부를 살펴보겠다.

### 3. ad hoc 가설의 기존 개념화들에 대한 검토

ad hoc 가설 개념의 정밀한 규명/형식화를 둔 현대의 대표적 입장들은 크게 다음과 같이 나뉜다:

- I. ad hoc 가설에 대한 통계적 단순성에 입각한 견해
- II. ad hoc 가설에 대한 베イズ주의적 견해<sup>9)</sup>

9) 단, 이때 해당 분류에서 I과 II가 엄격히 상호배치되는 것은 아니다. 3.2절에서 소개될 AIC와 더불어 가설의 통계적 단순성을 따지는 대표적인 기준중 하나는 베イズ 정보 척도*Bayesian information criterion*이며 이는 베イズ주의가 채택하는 주관주의 확률해석에 기초하기 때문이다; 또한 Kieseppä(2001)은 위 두 척도 모두 베イズ주의에 입각하여 정초될 수 있다고 주장하기도 한다. 하지만 본

### Ⅲ. ad hoc 가설 개념 자체에 대한 회의적 견해:

1. ad hoc 가설 개념 자체가 일관되게 사용되지 않는다는 입장
2. ad hoc 가설 등 이론 선택 시 고려사항이 외부적(사회적) 요인에 의하여 결정된다는 입장

이러한 일련의 입장들을 비교함에 있어 본고에서는 크게 다음 두 가지 조건을 중심으로 각 견해들을 검토하고자 한다:

첫 번째 **조건 (ㄱ)**은 각 입장들이 ‘ad hoc 가설’에 관하여 **이론내재적 기준**을 제시하는지 여부이다. 이때 기준이 이론내재적이라 함은 ad hoc 가설 해당 여부를 가르는 척도가 해당 보조가설 및 이론에 대한 분석을 통하여 결정될 수 있다는 것을 뜻한다. 이에 해당하지 않는 예시들로는 이론외재적 기준을 제시하는 입장(예. 과학자 사회의 역학 관계에 기초한 입장) 및 명시적 기준 자체에 대해 회의적 태도를 취하는 입장 등을 들 수 있다. ‘ad hoc 가설’ 기준이 이론내재성을 만족시킨다는 것을 확인할 수 있을 경우, 이는 발견의 맥락 *context of discovery*과 정당화의 맥락 *context of justification*간의 구분을 비롯하여 과학사를 내적 역사로 재구성하는 것의 근거가 된다.

두 번째 **조건 (ㄴ)**은 각 입장들이 **관찰의 이론적재성**과 합치하는지 여부이다. 이론적재성이 이론 시험/선택의 맥락에서 지니는 핵심적인 함축은 다음과 같이 요약될 수 있다:

- 서로 배치되는 이론을 지니는 관찰자 T1과 T2가 동일한 객관적 대상 O를 관찰할 때 두 관찰자가 수용하는 관찰의 내용 E1과 E2는 서로 다르다.

이는 곧 이론 시험/선택 상황에서 참조되는 경험적 내용이 객관적으로 동일한 것이 아니라, 주관적 수용으로 인해 서로 차이가 난다는 점

---

고에서는 ‘ad hoc 가설’을 형식화하는 상이한 방식에 주목하고 있으므로, 각 입장을 따로따로 논하고자 한다.

을 의미한다. 본고의 전제 자체가 관찰의 이론적재성을 수용하는 것이므로, 해당 제약조건 (⊥)를 만족시키는 것은 각 입장들을 평가함에 있어 우선적인 사항이 된다.

따라서 본고에서는 다른 요건 및 가치들에 앞서 위 조건 (⊃), (⊥)을 만족시키는지 여부를 우선적으로 살펴보고자 한다.

### 3.1. ‘ad hoc 가설’ 개념에 대한 회의적 견해

상기 입장 분류 가운데 III에 속하는 견해들은 ad hoc 가설 개념 자체 혹은 그 이론-내재적 형식화에 대하여 부정적인 입장을 취하는 견해들이다. 앞에 소개된 쿤의 견해 역시 이 범주에 속한다고 볼 수 있다.

III.1에 속하는 입장으로는 그렉 뱌포드와 크리스토퍼 헌트의 입장을 들 수 있다: 이들은 공통적으로 ‘ad hoc 가설’ 개념 자체의 일관성에 대하여 회의적인 입장을 드러내며, 주로 포퍼의 견해로 대표되는 관찰가능성 기조(2절 서두의 기조 1)를 논박하는데 초점을 기울인다.

뱌포드(Bamford, 1999)는 ‘ad hoc 가설’의 정의인 ‘주어진 현상 밖에 설명하지 못하는 보조가설’은 존재하지 않는다고 주장한다: (설명항) → (피설명항)이라는 도식을 전제할 때, ‘ad hoc한 설명항(=가설)’이란 설명항이 흡사 피설명항의 동의반복이나 마찬가지로인 경우를 가리킨다. 하지만 양 공간에 아무런 실질적 내용상의 차이가 없는 것처럼 보여도, 설명항은 당대의 언어 맥락에서 (설명항상의 어휘들이 그와 연동된 새로운 경험적 함의를 지니게 됨에 따라) 얼마든지 별개의 현상들을 설명할 수 있다. 사뭇 의미 전체론에 입각한 것 같은 해당 논문의 결론은 곧 ad hoc 가설이 “환영chimera”에 지나지 않는다는 것이다.

헌트(Hunt, 2012)는 에테르 이론을 지키기 위해 도입되었던 피츠제럴드-로렌츠 수축(FLC)을 중심 사례로 두고 포퍼의 ad hoc 가설 개념화를 비롯한 여러 견해들을 검토한다; 과학사적 사례들에 비추어봤을 때 2절에서 이미 설명한 ‘원리적 관찰불가능성’에 대한 논박을 비롯하여 ad

hoc 가설의 다양한 개념화 방식들은 전부 비일관적이며, 결국 연구자 개인의 미적 감각에 달린 문제로 밖에 파악될 수 없다고 해당 논문은 주장한다.

반면 III.2는 ad hoc 가설을 비롯한 이론 시험/선택에 개입하는 요소들이 이론-내재적으로 판단되는 것이 아니라, 이론-외재적 상황에 의하여 결정된다는 입장이다. 데이빗 블루어(Bloor, 1991)의 지식사회학 기획에 입각한 이러한 접근 방식의 예시로는 개별 과학자가 특정 이론을 받아들임으로써 획득하는 효용 *utility*을 변수로 삼아 과학자 사회가 특정 이론을 받아들일지 여부에 관한 판별식을 세우고자 하는 연구를 들 수 있다(Brock and Durlauf, 1997).

III에 속하는 상기 견해들은 공통적으로 ‘ad hoc 가설’ 개념 자체, 혹은 그에 대한 이론-내재적 정의를 시도하는 것에 대하여 부정적인 입장을 취한다. 그러나 III.1에서 논박하는 ‘ad hoc 가설’ 개념이 관찰가능성을 기조로 한 입장에 국한된다는 점, III.2에선 이론-내재적 고려가 처음부터 언급이 이루어지지 않는다는 점 등을 미루어볼 때, 이들은 이론적재성을 전제하는 동시에 이론-내재적 개념화를 시도하는 본고의 기획과는 아예 다른 전제에서 출발하는 것으로 평가할 수 있다. 즉, 이는 상기한 바 제약기준 (7)에 어긋난다.

따라서 본고에서는 이러한 회의적 견해들에 관한 평가는 유예하고, ‘ad hoc 가설’ 개념화를 적극적으로 시도하는 입장들을 중점적으로 살펴보고자 한다.

### 3.2. ‘ad hoc 가설’ 개념에 대한 긍정적 견해

‘ad hoc 가설’ 개념화를 적극적으로 시도하는 시도들은 대개 확률론적 맥락에서 이루어진다. 경험과학의 여러 기제를 양적으로 분석하는데 쓰이는 많은 귀납논리적 접근이 확률 개념을 적극적으로 사용하기 때문이다<sup>10)</sup>. 하지만 본고에서는 이런 방식으로 ad hoc 가설을 해명하고자

10) 즉 어떤 확률 개념 해석을 택하는지에 따라 ad hoc 가설 개념에 대한 접근

하는 대표적인 견해들(I,II) 모두 본고에서 제안하는 바 ‘ad hoc 가설’의 형식화 조건(( $\neg$ ),( $\neg$ ))에 부합하지 못한다는 점을 보이고자 한다<sup>11)</sup>.

### 3.2.1. 통계적 단순성에 입각한 ‘ad hoc 가설’ 개념화

이와 같은 적극적 입장에 속하는 대표적 접근 방식으로는 이론의 통계적 단순성에 입각하여 ad hoc 가설 개념을 해명하려는 조류(I)가 있다. 이는 통계학적 관점에서 가설을 설계시, 주어진 관찰 자료들에 가설이 과도적합 $over-fitting$ 하게 되는 것을 방지하는 문제와 맥락을 같이 한다; 관찰자료가 오류 $noise$ 를 내재할 수 있음을 전제할 때, 가설은 관찰자료에 과도적합하게 됨에 따라 불필요한 매개변수의 증가로 인해 그 복잡도가 높아지고, 나아가 차후의 관찰자료에 대한 예측력을 상실하는 것으로 알려져있기 때문이다(Hawkins, 2003; Baker, 2013).

이러한 방식으로 ad hoc 가설 개념을 정의하는 대표적 시도로는 아

---

역시 달라질 수 있다: 예컨대 II(‘베이즈주의’)는 빈도주의 같은 비-주관주의적 확률 해석을 허용하지 않는 반면, I에 해당하는 견해(e.g. AIC)는 앞의 각주에서 언급된 것처럼 보다 다양한 해석을 용인하는 것처럼 보인다. 본고에서는 II에 관한 논의(3.2.2절)에서 이런 확률 해석으로부터 비롯된 특징을 주로 참조하겠다.

- 11) 주류 견해에 속하지는 않지만, Scott(2012)은 논리/확률적 관점에서 ad hoc 가설을 ‘기존 이론에 견주었을 때 임의적인 가설’로 정의하는 개념화를 시도한다: 보조가설 H에 관하여  $T = \{(\text{이론의 핵심 명제들}), (H \text{와 충돌하지 않는 보조가설들}), (\text{이상현상})\}$ 이 있을 때, Scott은 T가 H를 논리적으로도/확률적으로도 함축하지 않는 경우 오직 그 경우에 H는 ad hoc 가설이며, 그 반대로 T가 H를 논리적으로/확률적으로 함축하는 경우 오직 그 경우에 H는 비-ad hoc 가설이라고 주장한다. 하지만 이러한 주장과는 달리, ad hoc 가설은 비-ad hoc 가설에 비해 기존 이론과 오히려 더 높은 논리적/확률적 유관성을 지닌다; ad hoc 가설은 기존 이론의 핵심적 명제들을 보존하는 목적하에서 그와 합치하는 형식으로 제시되기 때문이다. 예컨대  $T = (\text{르네상스 당시의 천문학 및 당대의 이상현상})$ ,  $H_1 = (\text{천동설-주전원 가설})$ ,  $H_2 = (\text{지동설-케플러 가설})$ 이라고 할 때에  $p(H_1|T) > p(H_2|T)$ 이라고 보는 것은 자연스럽다. Scott의 이러한 문제점은 T의 원소 가운데 기존 이론과 이상현상이 논리적/확률적으로 서로 길항하며, 또한 ad hoc 가설의 판단시점이 이론들 간의 경합이 이루어질 때라는 점을 간과함에 따라 발생한 것으로 보인다.

카이케 정보 척도 *Akaike information criterion* (이하 “AIC”)를 이론 시험/선택의 기준으로 취하는 말콤 포스터와 엘리엇 소버의 견해를 들 수 있다(Forster and Sober, 1994). 통계적 관점에서 임의의 과학적 이론/가설은 좌표계상의 한 곡선으로 투영될 수 있다. 이때 각 이론에 대응하는 곡선들은 그 차수에 따라 일련의 곡선족들로 나뉘어 분류된다; 예컨대 2차 곡선으로 투영되는 모든 이론들은 하나의 곡선족에 속한다. AIC의 정의에 입각하여, 포스터와 소버는 곡선족  $M$ 의 예측적 정확도 *predictive accuracy*를 다음과 같이 정의한다(10):

- (아카이케 정리): ( $M$ 의 예측적 정확도)  $\asymp (1/N) \times (\log\text{-likelihood}(L(M)) - k)$

위 식에서 ‘ $N$ ’은 전체 관찰자료의 개수, ‘ $L(M)$ ’은  $M$ 의 원소 가운데 관찰자료들에 가장 부합하는 곡선, ‘ $\log\text{-likelihood}(L(M))$ ’은 해당 곡선과 관찰자료 간의 우도 *likelihood*를 나타내는 함수, ‘ $k$ ’는  $M$ 에 속하는 곡선들의 차수를 뜻한다. 즉 ‘ $\log\text{-likelihood}(L(M))$ ’은 주어진  $M$ 의 원소 가운데 최적의 곡선이 주어진 관찰자료와 부합하는 정도에 상응하며, 그에 반해 ‘ $k$ ’는 해당 가설의 이론적/통계적 복잡도에 상응한다.

이때 주목할 만한 점은 예측적 정확도를 결정함에 있어서 관찰자료와 이론간의 부합도와 이론의 단순성이 서로 길항하는 것으로 형식화된다는 점이다; 즉 위 형식화는 어떤 이론이 이론적 단순성을 도외시킨 채 관찰 자료와의 부합도만을 제고함에 따라 과도적합 문제를 야기함에도 불구하고 여전히 ‘예측적으로 정확하다’는 평가를 받는 사태를 미연에 방지한다.

포스터와 소버는 위 형식화에 입각하여 라카토슈의 ad hoc 가설에 관한 모형을 보완하고자 한다: 이들은 라카토슈의 모형에서 ‘(미래에) 예측될 수 있는 관찰자료’를 기준으로 ad hoc 가설을 판별하는 것은 “예측적 성공”과 “예측적 힘”을 혼동하는데서 비롯된다고 주장한다; 라카토슈의 모형은 예측적 힘의 유무를 판별하는 데는 도움을 주지만, 이는 반드시 예측의 성공을 보장하지는 않기 때문이다(17).

따라서 이들은 연구 프로그램에 수정이 가해질 때 그로 인해 초래되는 이론-관찰 간 부합도의 상승분이 이론 단순성의 하락분을 상쇄하지

못하는 경우, 해당 연구 프로그램은 퇴행적이라고 주장한다(17). 이는 상기 도식을 차용할 경우, 다음과 같이 형식화될 수 있다:

- $M$ 에 보조가설  $A$ 가 추가되어  $M'$ 가 될 때,  $\log\text{-likelihood}(L(M')) > (M' \text{의 차수})$ 일 경우, 오직 그 경우에만  $A$ 는 ad hoc 가설이다.

그러나 이론 시험/선택 문제에 관한 유력한 견해로 받아들여짐에도 불구하고, 해당 견해는 본고의 기획과는 합치되지 않는다. 이에 관해서는 크게 두 가지 이유를 들 수 있다.

그 첫 번째 이유는 포스터와 소비의 형식화가 이론적재성을 고려하지 않는다는 점에서 제약조건 ( $\perp$ )을 위반한다는 점이다. 물론 이는 관찰자료가 오류를 포함하고 있을 가능성을 전제한다. 그러나 해당 형식화에서 각 이론 모형의 비교 척도가 되는 예측적 정확도의 정의를 살펴볼 때, 관찰자료에 관한 곡선족의 원소 가운데 최적 곡선( $L(M)$ )을 고르는 과정에서 관찰자료 집합 자체는 각 곡선들에 걸쳐 두루 유일한 것으로 상정된다(9). 각 곡선이 곧 별개의 과학적 이론에 상응하는 점을 고려할 때, 이는 곧 어떤 이론을 취하는지와 무관하게 유일한 관찰자료만이 주어진다는 것을 뜻한다; 이는 제약조건 ( $\perp$ )을 위반하는 것이다.

혹자는 AIC 척도의 기본 골격을 유지하면서도 위 문제점을 보수할 수 있다는 반론을 펼칠 수도 있다. 예컨대 관찰자료를 동일/유일한 것으로 상정하는 것이 아니라, 각 곡선에 따라 관찰자료 집합 역시 달라진다는 수정된 가설을 취할 경우, 원 저자의 기획과는 멀어짐에도 불구하고 위 제약조건 ( $\perp$ )은 만족될 수 있다. 그러나 이런 ad hoc한 방편에도 불구하고, 해당 견해는 이론적 단순성의 조작적 정의의 측면에서 보다 근본적인 두 번째 문제를 지니는 것으로 보인다.

위 AIC 견해를 비롯하여 통계학적 관점에서 ad hoc 가설 문제에 접근하는 방식 일반은 이론내 변환가능한 매개변수 *adjustable parameter*의 개수 혹은 이론이 나타내는 곡선식의 차수가 곧 이론의 복잡도를 나타낸다는 조작적 정의를 취한다. 그러나 해럴드 제프리스(Jeffreys, 1998[1961])에 의해 최초로 제안되었다고 알려진 해당 정의가 본고에서



의 ‘ad hoc 가설’ 개념에 부합하는지 여부에 대해선 의문의 여지가 있다.

가설 곡선의 차수와 그 이론적 복잡성을 등치시키는 조작적 정의의 대표적인 옹호 논변은 다음과 같다: 정수  $n < m$ 일 경우,  $n$ 차식의 집합은  $m$ 차식 집합 가운데  $(n+1)$ 차  $\dots$   $m$ 차항의 계수가 0인 진부분집합에 해당한다. 즉 실재하는 법칙이  $n$ 차식 곡선에 대응하는 경우 이는  $m$ 차식 집합에도 포함되지만 그 역은 성립하지 않으므로,  $n$ 차식으로 상정되는 이론의 외연은  $m$ 차식으로 상정되는 이론의 외연보다 작다. 그러므로  $n$ 차식 이론을 제안하는 것은  $m$ 차식 이론보다 더 제한적/구체적 주장이다. 따라서  $n$ 차식 이론은  $m$ 차식 이론보다 더 단순하다는 결론이 도출된다.

단 이때 마지막 도출 단계에서 사용되는 “제한적/협소한 내용을 주장하는 이론이 곧 단순한 이론이다”라는 가정, 즉 대우를 취할 경우 “복잡한 이론은 비제한적/광범위한 내용을 주장하는 이론이다”라는 가정은 합당한가? 이를 옹호하기 위한 논거로는 먼저 상기한 바처럼 가설이 제한적일 수록 과도적합의 위험성이 줄어든다는 직관이 있을 수 있으며(Forster and Sober, 1994, 8-9), 또한 이론의 정보성(*informativeness*)이 이론의 단순성과 밀접한 연관을 맺고 있다는 견해도 제시될 수 있다(Kieseppä, 2001). 그러나 과학사적 사례에 비추어볼 때 이러한 가정이 타당한지 여부에 관해선 재론의 여지가 있다.

예컨대 복잡한 이론과 단순한 이론의 대치로 인하여 보다 단순한 이론이 받아들여지는 전형적 예시인 르네상스 당시 천동설과 지동설간의 경합을 검토해보자<sup>12)</sup>. 이때 천동설은 보다 복잡한 주전원 보조가설 등이 부가되었으므로 지동설에 비해 더 복잡도가 높다고 말할 수 있지만, 천동설이 지동설에 비해 보다 비제한적/광범위한 내용을 주장하는 이론이라고 주장할 수는 없는 것 같다. 왜냐하면 천동설 또한 지동설이 그러하듯 분명히 일정한 시험적 귀결(*test implication*)을 지니고 있으며,

12) 그러나 코페르니쿠스 등 당대 천문학자들이 인식한 바와 달리, 실제 주전원 가설 등 보조가설들이 사용되던 실태를 고려할 때 당대의 지동설이 반드시 천동설보다 단순하지는 않았으리라는 주장도 있다(cf. Fitzpatrick, 1.b.ii). 단 본고에서는 어디까지나 단순한/복잡한 이론의 전형으로서 양 이론을 제시하는 것이므로, 보다 전통적인 학설을 따른다.

오히려 보조가설이 복잡하게 부가된 천동설이야말로 연연지가 더 많은 점에서 보다 제한적인 시험적 귀결을 지닐 개연성마저 있기 때문이다. 아울러 포퍼는 제프리스의 ‘단순성’에 관한 견해를 비판하며, 실제 수식의 측면에 있어서 마저도 전통적 천동설에서 취하는 원형 주전원들로 이루어진 궤도가 케플러의 타원형 궤도보다 변환가능한 매개변수의 수가 적다는 점을 지적하기도 한다(Popper, 2002[1959]; cf. Fitzpatrick, 4.a.iii). 따라서 AIC 이론을 비롯한 통계적 관점에서 시도되는 이론의 단순성에 대한 조작적 정의는 전적으로 일반화될 수는 없고, 그러므로 ad hoc 가설의 개념화에 관해서도 유효하지 않은 것으로 보인다.

정리하자면 AIC 등 변환가능한 매개변수에 입각한 통계적 단순성 개념을 통하여 ad hoc 가설을 개념화하는 기획은 먼저 제약조건(⊥)을 어기며, 나아가 “복잡도(단순성)” 개념어의 적용에 있어서 과학사적인 용례와는 차이점을 나타내는 것으로 보인다.

이러한 문제는 당초 ad hoc 가설의 개념화 기조 가운데 2안에 해당하는 ‘이론 복잡도’를 판단하는 것이 어려운 과제이며, 어쩌면 “복잡도(단순성)”은 그 자체로 중의적인 개념어일 가능성을 시사한다. 이에 관하여 염두에 둘 점은 ad hoc 가설의 개념화 방안으로서의 ‘이론 복잡도’가 이론 시험/판단 절차에서 수행하는 역할이다. 프톨레마이오스 천문학의 복잡성을 두고 13세기의 Alfonso X가 “만약 신이 나에게 세계 창조에 관하여 조언을 구하였더라면, 꽤 좋은 조언을 해줄 수 있었을 텐데”라고 신랄하게 논평하였듯(Kuhn, 1970, 69), ad hoc 가설의 잠재적 증거로서 이론의 복잡도는 이론으로부터의 이반을 야기할 만큼 연구자에게 강력한 동인을 제공해야만 한다.

### 3.2.2. 베이즈주의에 입각한 ‘ad hoc 가설’ 개념화(II)

ad hoc 가설 해명을 위한 두 번째 방식(II)으로 제안된 **베이즈주의 Bayesianism**는 주관주의 *subjectivism* 확률해석의 일종 혹은 그에 입각한 과학적 입증 이론을 가리킨다. 주관주의 확률해석의 핵심적 특징은 확

를 ' $p(X)$ '를 ' $X$ 에 대한 믿음의 정도'로 해석한다는 점, 그리고 베이즈 규칙상 사전확률 *a priori probability*의 적극적 사용을 인정한다는 점이다.

베이즈주의는 이러한 확률해석에 입각하여 관찰  $e$ 에 입각한 가설  $h$ 의 입증 및 반입증을 설명한다.  $h$ 에 대한 신뢰도('사전확률  $p(h)$ '),  $e$ 에 대한 신뢰도('근거확률  $p(e)$ '), 그리고  $h$ 를 전제시  $e$ 가 도출될 확률('우도  $p(e|h)$ ')을 확보할 수 있을 때, 베이즈 규칙에 따라  $e$ 를 반영한  $h$ 의 사후확률 *a posterior probability* ( $p(h|e) = p(e|h) \times p(h)/p(e)$ )를 계산할 수 있다. 베이즈주의에선 과학 이론의 발전을 이렇듯 가설  $h$ 의 확률이 바뀌어가는 과정으로 설명한다.

베이즈주의의 도식상 'ad hoc 가설' 분류 기준은 보조가설  $a$ 에 관하여 적용되며,  $a$ 는  $h$ 와 더불어 사전확률을 결정하는 확률변수로 사용된다; 예컨대  $h$  = '지구는 우주의 중심이다.',  $a$  = '태양/행성은 주전원 궤도에 따라 원형 궤도로 움직인다'에 관하여 확률값  $p(h)$ ,  $p(a)$  및 그 결합분포  $p(h,a)$  등을 확인할 수 있다는 것이 베이즈주의의 골자이다. 이때 ad hoc 가설  $a$ 는 이론  $h$ 에 대한 반증 사례  $e$ 가 관찰됨에도 불구하고 ( $p(e|h)=0$ ),  $a$ 로 인하여  $h$ 에 대한 신뢰도( $p(h)$ )가 유지되는 경우에 해당한다.

베이즈주의자들은 이에 관하여  $p(h)$ 와  $p(a)$ 의 상대적인 비율이 반대 사례의 반입증 정도와 반비례한다는 사실에 주목한다; 즉 가설에 대한 신뢰도가 보조가설에 대한 신뢰도보다 현격히 높을 경우 반입증 사례가 관찰될 때에도 가설에 대한 신뢰도에 대한 영향은 미미하며, 반대로 보조가설에 대한 신뢰도가 현격히 높은 경우 반입증의 효과는 가설에 집중된다. 예컨대 가설  $h$ 에 대한 신뢰도는 높고( $p(h)=0.9$ ) 보조가설  $a$ 에 대한 신뢰도는 상대적으로 낮은 경우( $p(a)=0.6$ ), 반증 사례가 출현 시( $p(e|h,a)=0$ ,  $p(e|\sim h,a)=0.01$ ,  $p(e|\sim h,\sim a)=0.01$ ,  $p(e|h,\sim a)=0.02$ ) 가설 신뢰도에 대한 영향은 미미한 반면( $p(h|e)-p(h) = -0.022$ ) 보조가설에 대한 신뢰도는 현격하게 떨어진다( $p(a|e)-p(a) = -0.527$ ).

즉 천왕성 도입 가설 등 '성공적'인 보고가설 예시들은 이론 신뢰도가 보조가설 신뢰도에 비해 훨씬 더 높았던 경우로 간주되는 반면, 보조가설이 "ad hoc하다"는 이유로 받아들여지지 않는 경우는 보조가설

신뢰도가 기존 이론 신뢰도에 훨씬 떨어지는 경우로 분류된다(Strevens, 2001). 이렇듯 베이즈주의자들은 프라우트의 원자량 정수치 가설 등 과학사의 예시를 위 도식을 통하여 설명함으로써, 확률론의 공리에 입각한 베이즈주의 모형이 실제 과학의 발전 과정을 잘 설명해준다고 주장한다(Howson and Urbach, 2006).

이러한 베이즈주의적 접근은 “이론의 반증 방식에 대한 상세한 고려”(Strevens, 2001, 39) 없이도 보조가설의 ad hoc 가설 해당 여부 및 그에 따른 이론의 구체 여부를 설명할 수 있다는 장점을 지닌다. 그러나 이러한 특징은 다른 한편으로 베이즈주의가 한계를 내포하고 있다는 점을 알려주는 단서이기도 하다. 베이즈주의의 상기 특징은 베이즈주의가 결국 이론의 내용이 아닌, 이론에 대한 과학자(집단)의 신뢰도라는 주관적 척도에 기초한다는 점을 시사하기 때문이다.

예컨대 ‘ $T$ =프톨레마이오스 천문학’라고 정의하자. 주관주의 확률해석을 전제할 때, 서기 4세기 무렵의  $p(T)$ 의 값과 현대의  $p(T)$ 의 값은 명백하게 차이가 난다. 이렇듯 가설/보조가설에 대한 신뢰도는 시대, 집단, 상황에 따라 가변적이므로, 이에 입각한 베이즈주의적 ad hoc 가설의 판단 기준 또한 시대, 집단, 상황에 따라 상대적일 수밖에 없다. 이는 확률해석 및 가설검증의 측면에서 주관주의/베이즈주의가 사전확률을 수용한다는 점 때문에 비판을 받는 것과 맥락을 같이 한다.

위 예시에서 나타나듯 베이즈주의는 당대 과학자 집단이 지니는바 이론에 대한 인식의 모사를 통하여 관찰의 축적에 따른 과학 이론의 검증/선택 과정을 설명할 수 있다. 따라서 이는 과학철학의 목표중 하나인 과학사에 대한 합리적 재구성의 가능성을 열어놓는 방법론이라는 점에서 긍정적인 면모를 지닌다. 그럼에도 불구하고 이러한 재구성은 어디까지나 과거의 과학사에 대해 적용될 수 있을 뿐이라는 점에서 사후적이다. 상기 예시에서 나타나듯 이론에 대한 당대의 인식은 시대, 장소, 상황에 따라 상대적이기 때문이다. 그러므로 이러한 방법론은 ‘ad hoc 가설’의 판단 기준을 이론 자체가 아닌 그에 대한 당대의 믿음이라는 외적 요소에 둔다는 점에서 제약 조건 (7)에 어긋난다.

다른 한편 베이즈주의는 과학자들의 믿음에 초점을 기울인다는 점에서 쿤으로 대표되는 비-경험주의적 과학철학자의 이론을 해석하는데 사용되기도 한다(Salmon, 1990; Earman, 1992, ch.8; Chen, 2006). 특히 이러한 입장들은 쿤의 이론의 합리적 재구성을 목표한다는 공통된 목표를 지니는 것으로 보인다.

베이즈주의에선 앞에서 언급한 바처럼 신뢰도가 높은 이론과 상대적으로 신뢰도가 낮은 보조가설이 주어질 경우, 반증사례가 관찰될 때에도 이론의 신뢰도 하락은 미미하다. 하지만 반증사례가 지속적으로 축적됨에 따라 이론의 신뢰도가 줄어듦으로써, 이론은 종국엔 과학계에서의 공인된 지위를 상실한다(Howson and Urbach, 2006, 113). 이는 독점적 위치를 차지하던 정상과학 패러다임이 이상현상의 축적에 따라 점차 위기에 봉착하게 되고 결국 혁명을 통하여 패러다임이 전환되는 쿤의 모형과 흡사하다. 상기 학자들은 이런 유사성에 주목하여 베이즈주의가 쿤 이론의 함의를 형식화할 수 있다고 주장한다.

하지만 레프테리스 파르마키스(Farmakis, 2008)는 쿤의 이론과 베이즈주의가 합치하지 않는 부분을 제시함으로써 위와 같은 시도를 반박한다: 해당 논문은 쿤의 ‘공약불가능성’에 대한 해석에 있어 별개 패러다임간 번역의 가능성을 아예 부정하는 강한 입장과 별개 패러다임간의 부분적 소통을 인정하는 약한 입장을 구분한다; 이때 후자는 “어느 한 과학적 전통에 소속된 베이즈주의적 주체라 할지라도 상대방의 (주장) 내용에 대해 부분적으로 잘 정의된 믿음을 가질 수 있게 한다”(49)는 측면에서 각 패러다임에 대한 신뢰도를 베이즈 규칙상의 사전확률로서 정의하는 것을 용인하는 반면, 전자는 아예 상대의 입장을 헤아릴 수 없다는 점에서 경합 패러다임을 사전확률로서 정의하는 것을 차단한다. 즉, 공약불가능성에 대한 강한 입장의 경우, 베이즈주의적 이론 시험/판단에서 요구되는 사전확률에 대한 정의 시도와 충돌한다.

하지만 이런 차이에도 불구하고 공약불가능성에 대한 두 입장은 공통적인 문제점을 지닌다: “베이즈주의가 문제를 세움에 있어 요구하는 바 독특한 방식은 동일한 근거evidence에 입각한 두 이론의 사후확률을 비교하는 것이다”. 하지만 파르마키스는 ‘이론 내 핵심적 명제들은

“경험 그 자체를 형성하는 제거불가능한 역할”을 수행하는 선험적 종합 명제에 해당한다”는 요지의 쿤의 진술을 전거로 들며, 공약불가능성에 관한 두 해석 모두 “한 이론에서는 유효하다 여겨지는 여러 중요한 [관찰]근거들이 다른 이론에서는 무의미하다”는 결론을 함축할 수밖에 없다고 주장한다. 즉 “공약불가능한 이론  $T_1 \cdots T_N$ 에 관하여  $T_i$ 를 지지하는 [베이즈주의적] 주체에겐 우도  $p(E|T_i, B)$  ( $i = 2 \cdots N$ )가 잘 정의될 수 없고 [...] 따라서 [베이즈주의적 이론 경합의 척도인] 사후확률 또한 정의될 수 없다”. 파르마키스는 이렇듯 “동일한 근거에 입각하여 [사후확률을] 갱신하는 것은 어떤 베이즈주의적 입장이라 할지라도 필수불가결하므로, 상기 결과는 쿤과 베이즈를 화해시키려는 시도를 치명적으로 무력화시킨다”고 논평한다(50-51)<sup>13)</sup>.

베이즈주의적 관점에 입각하여 쿤으로 대표되는 역사주의/비-경험주의적 견해를 설명하고자 하는 기획이 실패한다는 파르마키스의 평가가 만약 옳다면, 이는 ‘ad hoc 가설’의 형식화에 관한 베이즈주의적 견해 역시 동일한 문제점에 봉착함을 시사한다. 베이즈주의가 이론-종속적이지 않은 객관적 관찰 근거가 확보될 수 있음을 전제한다면, 이는 본고에서 제시한 바 조건 (⊥)에 부합하지 않기 때문이다. 이러한 문제는 비단 쿤 뿐만 아니라, 이론적재성의 요소를 수용하는 견해들 전반에 일반화될 수 있을 것으로 보인다.

정리하자면 베이즈주의는 그 이론적 강력함에도 불구하고, 상기된 바처럼 제약조건 (⊃), (⊥)을 모두 만족시키지 못하기 때문에 본고에서

13) Farmakis(2008)은 “패러다임-종속적인 방식의 베이즈주의”, 즉 객관적 관찰 근거를 포기한 채 “각자가 보는대로(“as they see it”) 근거를 확보하여 사후확률을 갱신하는 방식의 가능성을 언급한다. 해당 논문은 이러한 기획이 “이론 비교에 있어 근거가 이론에 미치는 영향을 측정하고자 고안된 바 베이즈주의의 근본적인 목표를 부정한다”고 반박한다(52). 본고는 이러한 지적에 동의하며, 나아가 이런 패러다임-종속적 베이즈주의 또한 제약조건 (⊃)을 만족시키지 못한다는 점을 재차 강조하고자 한다. 하지만 이렇듯 ‘각자가 보는대로’의 관점에서 이론 시험/평가를 시도하는 관점 자체만은 4절에서 소개될 본고의 기획과 맥락을 같이 한다.

요구하는 ‘ad hoc 가설’에 대한 형식화 기준을 충족시키지 못한다고 평가할 수 있다.

#### 4. ‘ad hoc 가설’에 대한 대안적 형식화: ‘이론의 시간 복잡도’

본고에서는 지금까지 ‘ad hoc 가설’ 형식화에 관한 기존의 여러 논의들이 (ㄱ) 기준의 이론-내재성 (ㄴ) 관찰의 이론적재성을 만족시키지 못한다는 점을 밝혔다. 본고에선 그 대안으로 2절에서 제시한 ‘ad hoc 가설’ 형식화의 두가지 기조 가운데 ‘2. 이론의 복잡도(단순성)’에 초점을 두는 방식을 제안하고자 한다. 그러나 단순성을 미적 가치로 간주하는 쿤의 견해(2절)나 AIC 같은 통계적 방식에 입각한 견해(3.2.1절)에서 나타나듯, ‘이론적 복잡도(단순성)’는 중의적 개념어이며 보다 자세한 개념적 규정이 요구된다. 따라서 본고에서는 ‘이론의 시간 복잡도’에 의거한 구체적 형식화를 시도하고, 이러한 방식이 위 두 제약조건을 만족시킨다는 것을 보이고자 한다.

##### 4.1. ‘이론의 시간 복잡도’ 개념 도입 및 제약조건 검토

전산학에서 **시간 복잡도** *time complexity*는 계산 복잡도 *computational complexity*의 일종으로, 일반적으로 특정 튜링 기계가 주어진 입력 내용을 처리하기까지 소요되는 (튜링 기계 머리의) 이동 횟수로 정의된다. 일반적인 컴퓨터 언어는 복수-테이프 튜링기계상의 알고리즘과 흡사하므로, 계산 이론에서 시간 복잡도를 측정하는 기법은 실제 디지털 컴퓨터 프로그램의 소요시간을 측정하는데도 응용된다(Linz, 2001, 344-9). 일반적인 컴퓨터 프로그램 알고리즘의 소요시간을 측정할 때, 튜링 기계의 이동 횟수에 상응하는 것은 기초 연산 *primitive operation*의 수행 횟수이다. ‘알고리즘의 효율성’ 검사는 길이  $n$ 에 해당하는 임의의 입력 내용이 주어졌을 때, 해당 알고리즘의 점근적 소요시간 *asymptotic running time*에 대한 측정을 목표로 한다(Cormen et al., 2009, 25).

이른바 **처치-튜링 명제 Church-Turing thesis**는 어떤 알고리즘에 관한 튜링 기계의 존재가능성과 그 알고리즘에 관한 기계적/효과적 계산가능성을 동치라고 주장한다(Linz, 2001, 244-6; Copeland, 2008). 이는 형식화될 수 있는 이론체계의 경우 그에 해당하는 튜링 기계가 만들어질 수 있으며, 임의의 입력에 관해 그 결과값을 결정하기 위한 시간 복잡도를 측정할 수 있음을 시사한다.

경험과학 이론의 경우, 상기한 바 ‘기계적/효과적 계산가능성’에 상응하는 것은 ‘임의의 (가능한) 현상이 주어진 이론에 합치하는지 여부에 대한 철학적 판단 가능성(이하 “합치 판단 가능성”)’이다; 이때 합치 여부란 가설-연역 모형의 경우 이론 내 명제 및 보조가설들로부터 경험 명제가 도출되는지 여부에 해당하며, 베이즈주의의 경우 우도( $p$ (현상|이론))의 값으로 나타난다. 만약 합치 판단 가능성이 성립할 경우, 경험과학 이론에서도 상기한 바처럼 그 시간 복잡도에 대한 측정이 정당화된다.

물론 가설-연역 모형에서 전제하는 것처럼 경험과학 이론이 형식적으로 엄밀한 논리 체계로 재구성될 수 있는지 여부에 관해서는 단언하기 힘들다. 그럼에도 불구하고 상기 여러 과학철학 이론들을 살펴볼 때, 관찰의 이론에 대한 합치 여부는 항상 판단가능한 것으로 전제됨을 확인할 수 있다: 가설-이론 모형을 취한 반증주의에서 ‘이론/가설로부터 관찰명제가 도출될 수 있는지 여부’에 따라 이론의 용인 *corroboration* 혹은 반증 여부를 결정하는 점, 베이즈주의에서의 사후 확률 갱신을 결정하는 필수 변항 중 하나가 우도라는 점, 심지어 쿤의 이론에서조차도 정상과학 내에서 퍼즐풀이 활동의 성공 여부에 따라 이상현상 해당 여부를 가리는 점 등이 그 대표적 예시들에 해당한다. 이러한 특징은 연구자에 의해 관찰된 바가 어떠한 방식으로든 과학 이론에 환류 *feedback*된다고 전제하는 한 과학철학적 견해 일반이 공유하고 있는 것으로 보이며<sup>14)</sup>, 실제 과학적 활동 또한 그러한 것으로 여겨

14) 이렇듯 합치 여부에 대한 판단이 언제나 유한 시간 내에 이루어질 수 있다는



진다. 이러한 점들을 미루어볼 때, 합치 판단 가능성은 충분히 성립하는 것으로 보인다.

물론 과학적 이론에서 합치 여부에 대한 판단 가능성이 보장된다는 것을 보이는 것만으로는 임의의 관찰 자료에 대한 시간 복잡도 측정이 정당화될 수는 없다; 처치-튜링 명제에서 언급한 “기계적/효과적”, 혹은 앞선 정의에서 언급한 “절차적”이어야한다는 요건이 만족되어야하기 때문이다.

흔히 지적되듯 “기계적/효과적”이라는 요건을 명시적으로 정의하려는 시도는 대개 순환적이라는 문제를 지니지만(Linz, 2001, 245), 임의의 방법/절차 M이 “기계적/효과적”이기 위한 조건으로 다음과 같은 요건들을 잠정적으로 제시해볼 수 있다(Copeland, 2008):

- i. M은 유한한 수의 명시적 지침*exact instruction*으로 이루어져야 한다 (각 지침은 유한한 수의 기호를 통하여 표현되어야 한다).
- ii. 오류 없이 수행되는 한, M은 유한한 단계 내에 원하는 결과를 도출시킬 수 있다.
- iii. M은 (실제 혹은 원리적으로) 필기도구를 제외한 일체의 기계 장비 없이 사람에게 의하여 수행될 수 있어야한다.
- iv. M은 수행하는 사람에 대해 특별한 통찰력*insight* 혹은 독창성*ingenuity*을 요구하지 않는다.

---

점은 과학적 이론에 의거한 관찰의 합치 여부 판단 알고리즘이 튜링-계산가 능할 뿐 아니라, 오히려 더 강력한 규정임을 함축한다. 튜링 기계가 있다는 것은 ‘올바른 입력’이 이루어졌음을, 즉 관찰이 이론에 합치한다는 것만을 보장할 뿐, 합치하지 않는다는 판단을 유한시간 내에 보장하지는 않기 때문이다(cf. Linz, 2001, 281).

다만 본고에서는 경험과학적 이론들이 그에 해당하는 튜링 기계/알고리즘을 지닌다는 것을 기존 과학철학 견해 등의 경험적 근거에 의존하여 논하려는 것일 뿐이지, 이런 ‘경험과학적 이론’의 해당 여부를 결정하는 체계적 절차가 있음을 논하려는 것은 결코 아니다. 임의의 알고리즘과 입력에 관하여 해당 알고리즘이 입력 내용을 수용할지 아닐지를 메타적으로 판단하는 알고리즘(‘Halting Problem’)은 대표적인 계산불가능한 문제이기 때문이다.

이때 ‘과학적 이론에 대한 관찰의 합치 여부 판단’ 과정은 위 조건들에 부합하는가?

먼저 조건 i.의 경우, 문제가 되는 지점은 “명시적 지침”의 기준이 무엇이나는 점이다. 이때 요구되는 바 “명시성”이 형식적 규정을 의미할 경우, 과학 이론을 구성하는 요소들은 해당 조건을 엄격히 만족시킬 수는 없는 것으로 보인다; 설령 과학 이론에서 수식으로 구성된 부분이 ZFC-공리체계에 의거하여 엄밀하게 정의된 것으로 간주될 지라도, 여러 과학 이론은 많은 부분이 자연어로 구성되어 있는 것으로 보이기 때문이다.

하지만 이때 “명시적 지침”의 실질적 함축이 절차상 혼동되지 않을 만한 표준성을 확보하는데 있다고 한다면, 경험과학의 이론 또한 위 조건을 만족시킬 수 있을 것으로 보인다. 가설-이론 모형 같은 논리 경험주의적 입장을 취한다면 물론이거니와, 쿤의 ‘패러다임’ 정의에서도 나타나듯 이론은 과학 공동체 내에서 어떠한 설명/퍼즐풀이가 올바른 연구 수행으로 받아들여질 것인지를 결정하는 표준적 역할을 수행하기 때문이다(Kuhn, 1970, 184). 학계에서의 상호 심사 *peer review*는 이렇듯 과학 공동체에서 공유되는 이론적 절차의 올바른 수행을 감독하는 절차이며, 이러한 표준적 절차의 준수는 차세대 학자들에 대한 과학적 훈련을 통하여 지속적으로 유지된다. 따라서 과학적 이론에 대한 관찰의 합치 여부를 판단하는 과정은 충분히 명시적인 지침에 의하여 이루어진다고 판단할 수 있다. 아울러 괄호에 명기된바 ‘유한한 수의 기호’ 조건은 수식뿐 아니라 대부분의 자연어 역시 만족시킨다는 점을 고려할 때, “명시적 지침”을 엄격한 형식적 의미로 해석하지 않는 한 ‘과학적 이론에 대한 관찰의 합치 여부 판단’은 조건 i.을 만족시킬 여지가 있는 것으로 보인다.

조건 ii.의 경우, 앞에서 과학 이론에서의 합치 판단 가능성을 보인 점에서 이미 만족된다. 과학 이론에 대한 합치 판단의 수행 단계가 무한한 경우, 합치 여부를 특정할 수 없기 때문에 이론에 대한 환류가 불가능하기 때문이다.

조건 iii. 및 iv.은 평가자 및 실제 평가 과정에 관한 규정들이다: 이

들 규정들은 예컨대 ‘필기도구만을 사용하는 경우엔 원리적으로 풀 수는 있지만 적어도 지구가 멸망할 때까지는 다 풀리지 않는 경우’, ‘현대 수학 및 입자물리학 연구에 입각하여 문제를 푸는데 “특별한 통찰력 혹은 독창성”이 반드시 필요한지 여부’ 등에 관련하여 모호한 면을 지니며, 실제 튜링-계산가능한 모든 알고리즘에 관해서도 전적으로 일반화될 수 있는지 의심스럽다. 다만 수학 및 논리학 등 조건 iii.와 iv.을 만족시킨다고 흔히 간주되는 학제들을 고려할 때, 과학적 이론에 의거한 관찰의 합치 여부 판단 알고리즘 역시 크게 어긋나지는 않을 것으로 여겨진다.

따라서 처치-튜링 명제를 비롯한 상기 사항들을 받아들일 경우, 과학적 이론이 있을 때, 임의의 관찰이 그 이론에 대하여 합치하는지 여부를 판단하는 알고리즘이 있으며, 곧 그 알고리즘에 대응하는 튜링 기계의 시간 복잡도를 측정할 수 있다는 결론이 도출된다.

본고에선 이러한 이론의 시간 복잡도야말로 과학 이론의 복잡도(단순성)를 파악하는 대안적 척도가 되며, 나아가 ‘ad hoc 가설’ 해당 여부를 판단할 수 있는 기준이 된다고 주장한다; 보다 자세히 말하자면 과학 이론에 상응하는 비결정론적 튜링기계를 사용하여 주어진 현상과의 합치 여부를 풀 때 소요되는 튜링 기계 머리의 이동 횟수가 곧 주어진 현상을 설명함에 있어 ad hoc한 정도를 판단하는 기준이 된다. 비결정론적 튜링 기계를 가정함으로써, 해당 형식화는 현상과의 합치 여부를 보이기 위한 최단 시간 복잡도를 특정할 수 있다(“Nondeterministic Turing Machine,”).<sup>15)</sup> <sup>16)</sup>

15) 이때 이론과 현상간의 합치 여부를 보이는 대개의 문제는 NP-완비 문제에 해당할 것으로 추측되므로, ad hoc 가설 판단의 척도가 되는 바 비결정론적 튜링 기계를 통해 구하는 시간 복잡도는 다항식 시간일 것으로 보인다. 자세한 내용은 4.2절을 참조하라.

16) 상기 서술은 기존 이론( $T_1$ )에 보조가설이 더해짐으로써( $+A$ ) 수정된 이론( $T_1+A = T_2$ )이 결과적으로 나타내는 시간 복잡도를 포착한다. 하지만 이는 보조가설 도입 과정 자체( $+A$ )에서 초래되는 이론 복잡도를 포함하지는 않는다. 이에 관해서는 5절을 참조하라.

본고에선 기존 견해들의 요소를 계승하여 ‘ad hoc 가설’ 해당 여부는 2가적 판단이 아닌 연속적인 정도의 문제라고 간주하고, ad hoc 가설에 대한 판단은 복수의 이론 및 보조가설들의 경합상황에서만 상대적으로 이루어질 수 있다고 간주한다. 이러한 견지는 2절에서 제시된 바 ‘ad hoc 가설’의 형식화 기조 가운데 ‘이론적 복잡도(단순성)’를 기준으로 삼는 기조의 연장선상에 있다. 따라서 본고에서 제시하고자 하는 ‘ad hoc 가설’의 조작적 정의는 형식화 기조 2.를 수정하여 다음과 같이 제시될 수 있다.

- 보조가설 A가 이론 T에 부가될 때, A와 그 이후에 도입되는 보조가설들로 인해 T의 시간 복잡도가 (경합하는 다른 이론 및 보조가설들에 비하여) 현저히 높을 경우, A는 ad hoc 가설이다.

해당 기준은 이론내재적이라는 점에서 제약조건 (7)을 만족시킨다: ad hoc 가설의 기준으로서 단순성을 제시하며 이를 주관적/미적인 척도로 간주하는 쿤의 이론이나, 가변적 척도인 이론에 대한 신뢰도를 기준으로 ad hoc 가설을 판단하는 베이즈주의(3.2.2절)와 달리, 시간 복잡도는 이론 자체에 대하여 매겨질 수 있기 때문이다.

특정 알고리즘의 시간 복잡도는 점근적 차원에서 객관적으로 매겨질 수 있다: 예컨대  $n$ 개의 자료들을 정렬하는 작업의 경우, 삽입정렬 *insertion sort* 알고리즘의 소요시간은 최악의 경우에도  $O(n^2)$ 인 반면, 병합정렬 *merge sort* 알고리즘의 소요시간은  $O(n \times \log n)$ 이다(Cormen et al., 2009, 12)<sup>17)</sup>. 이렇듯 시간 복잡도는 특정 컴퓨터 언어에 종속되지 않는다는 점에서 객관성을 지닌다.

앞에서 우린 과학 이론이 있을 경우, 그에 해당하는 알고리즘 및 튜링 기계가 있음을 보였다. 따라서 과학 이론에 상응하는 알고리즘 및 튜링 기계의 점근적 시간 복잡도는 곧 해당 과학 이론에 종속적이며, 이를 받아들이는 과학 공동체의 인식 등과 무관하다. 그러므로 그 점

17)  $O(f(n)) = \{g(n) \mid \text{어떤 } c > 0 \text{과 어떤 } n_0 > 0 \text{에 관하여 } n > n_0 \text{ 인 모든 } n \text{에 대해 } f(n) < c \times g(n) \text{이 성립한다}\}$

근적 시간 복잡도를 과학 이론의 복잡도로 간주할 경우, 이는 이론내 재적이다.

또한 해당 기준은 관찰의 이론적재성에 충돌하지 않는다는 점에서 제약조건 (L)을 만족시킨다: 통계적 단순성에 입각한 견해(3.2.1절)가 핵심적 요소로 이론과 관찰자료 간의 부합도를 요구하며, 베이즈주의가 이론-종속적이지 않은 객관적 근거를 전제하는 것과 달리, 시간 복잡도를 측정하는 방식은 이론적재성과 충돌하지 않는다. 시간 복잡도에 대한 측정은 객관적 현상이 인식 과정에서 이론에 의해 왜곡되진 아니건, 관찰자가 수용하는바 주관적인 내용을 입력받아 이를 이론에 맞춰 처리하고 그 소요시간을 출력하는 것으로 충분히 정합적이기 때문이다. 즉, 본 방식은 이론과 관찰내용간의 부합 정도를 따지는 전통적인 대응론과는 전적으로 독립적이다.

이러한 주장에 관한 예상되는 반론으로는 입력 자료의 크기가 이론에 의한 왜곡에 의하여 변할 수 있다는 주장이 있을 수 있다: 예컨대 크기  $n$ 에 해당하는 관찰 근거를 받아들인다고 가정하자. 만약 이론 T1을 믿을 경우 관찰 자료엔 왜곡이 가해져 그 크기가  $n+1$ 이 되는 반면, 이론 T2를 믿을 경우엔 크기가  $n-1$ 이 된다고 할 경우, 이때 시간 복잡도를 측정하는 것은 객관성을 상실하는 것처럼 보인다.

하지만 이는 시간복잡도의 의의를 잘못 파악한 것에서 비롯된 비판이다; 본 기획의 핵심은 복수의 이론 및 보조가설들이 경합하는 경우 동일한 입력에 관해 요구되는 시간 복잡도를 서로 비교하는 것이며, 그 값은 가변적이지 않기 때문이다.

3.2.2절에서 사전확률이 시대, 장소, 상황에 따라 가변적이라는 점을 들어 베이즈주의를 비판했던 사례를 견주어 검토하여 보자; 베이즈주의는 객관성을 결여한다는 이유로 본고에서 논박당한 대표적인 예시이므로, 혹자는 시간복잡도 또한 같은 문제에 처한다는 점에서 함께 거부되는 것이 공평하다고 주장할 수 있기 때문이다.

이때 비교의 대상이 되는 바, ‘시간 복잡도 기획’과 ‘베이즈주의’ 두

맥락에서 각각 구하고자 하는 목표값인 ‘시간 복잡도’와 ‘사후확률’이 어떻게 결정되는지를 살펴보자. 이때 이들이 독립 변인들로부터 결정되는 과정을 다음과 같이 정리해볼 수 있다(“ $\langle X_1, Y_1 \dots \rangle \rightarrow \langle X_2 \dots \rangle$ ”는 “ $\langle X_1, Y_1 \dots \rangle$ 가  $\langle X_2 \dots \rangle$ 를 결정하는 독립 변인(들)이다”로 새긴다):

• (왜곡을 전제한) 시간 복잡도 구하기:

$\langle$ 본래 관찰 근거의 크기, 관찰자가 믿는 이론 $\rangle \rightarrow \langle$ (왜곡된) 관찰 근거의 크기, 관찰자가 믿는 이론 $\rangle \rightarrow \langle$ 시간 복잡도 $\rangle$

• (왜곡을 전제한) 베이즈주의:

$\langle$ 이론, 시간/장소/상황 $\rangle \rightarrow \langle$ 사전확률, 우도, 근거확률 $\rangle \rightarrow \langle$ 사후확률 $\rangle$

각각의 경우에 '왜곡'을 야기하는 것은 '관찰자가 믿는 이론' 그리고 '시간/장소/상황'이다. 그런데 이때 왜곡/가변이 일어나지 않는 이상적 상황을 가정해보자; 즉 이는 시간 복잡도를 구하는 맥락에선 이론에 의한 왜곡이 일어나지 않아서 관찰 근거의 크기가 변하지 않는 상황이며, 베이즈주의의 맥락에선 시간/장소/상황이 사전확률에 영향을 미치지 않는 상황이다(예. 시공간을 초월한 신의 관점). 이때  $\langle$ 시간 복잡도, 사후확률 $\rangle$ 을 각각 결정하는 독립 변인은 어떻게 변할까?

• (이상적 상황에서) 시간 복잡도 구하기:

$\langle$ 본래 관찰 근거의 크기, 관찰자가 믿는 이론 $\rangle \rightarrow \langle$ 관찰 근거의 크기, 관찰자가 믿는 이론 $\rangle \rightarrow \langle$ 시간 복잡도 $\rangle$

• (이상적 상황에서) 베이즈주의:

$\langle$ 이론 $\rangle \rightarrow \langle$ 사전확률, 우도, 근거확률 $\rangle \rightarrow \langle$ 사후확률 $\rangle$

위 두 도식을 비교 시, 시간 복잡도를 구하는 과제에선 이상적 상황을 상정하는 경우에도 최초의 독립 변인 목록에 아무런 변화가 없는 반면, 베이즈주의의 경우 독립 변인 목록에서 '시간/장소/상황'이 탈락함을 알 수 있다. 이상적 상황에서 사후확률을 결정할 때 '시간/장소/상황'이 제외되는 것은 이것이 본래 고려 범위 바깥의 외적 변수였다는 점을 뜻한다. 반면 시간 복잡도를 결정하는 변인의 목록은 이상적 상황이나 왜곡된 상황이나 변하지 않는다; ‘관찰자가 믿는 이론’의 역할

은 자료의 왜곡을 발생시키는 부정적인 영향을 끼치는데 한정되는 것이 아니라, 그 자체로서 시간 복잡도를 계산하는데 핵심적이기 때문이다. 즉 왜곡이 발생시키는 요소가 있다한들 이는 조건 (7)에 부합하는 이론-내적 변인이며, 그런 이론-내적 변인이 고정되는 한 ‘시간 복잡도’는 일정하게 유지된다.

예컨대 관찰  $e$ 에 관하여 사후확률  $p(\text{점성술}|e)$ 를 구한다고 해보자; ‘해당 시점이 중세다’, ‘해당 시점이 현대다’ 등의 변인은 예기치 못하게  $p(\text{점성술}|e)$ 에 영향을 미치는 외재적인 변인들이다. 반면 시간 복잡도를 측정하는 경우, ‘관찰자가 믿는 이론’은 관찰 자료의 크기를 왜곡할 수도 있는 변인이지만 이는 여전히 이론내재적이다. 따라서 본래 고려하기로 되어있는 내재적 변인들만을 고려하는 한, 목표값인 시간 복잡도는 예기치 못했던 독립 변인에 의해 변할 여지가 없다.

즉 앞에서 베이즈주의가 겪었던 문제점과 달리, 시간복잡도 측정의 경우 목표값인 시간 복잡도의 값은 예기치 못한 상황에 의하여 변하지 않는다. 관찰 자료 크기가 왜곡되는 것은 전적으로 관찰자가 믿는 이론의 종속 변인이므로 이는 조건 (7)에 문제 없이 부합한다. 그러므로 우리가 관심을 갖는 목표치인 시간복잡도가 바뀔 염려는 없다<sup>18)</sup>.

18) 이처럼 시간복잡도 척도는 설령 ‘이론 적재성을 받아들이면 관찰 내용의 크기가 변하는게 가능하며 또한 그 변하는 정도를 파악할 수 없다’라는 가설이 참이라고 가정하더라도 아무런 문제를 빚지 않는다. 하지만 위 가설이 실제로 참인지 여부 또한 별도로 따져볼 문제다. 아래 두 사고 실험은 결정적인 논변이라고 하기엔 매우 막연하지만, 그럼에도 위 가설을 의심해볼만한 계기를 제공한다:

- (1) 감각적/심리적 이론적재성: 객관적 현상에 대응하여 물리적 자극의 감각기 수용 시점( $t$ )부터 관찰자의 의식에 감각소여가 현전할 때( $t'$ )까지 자기 동일성을 유지하는 ‘감각내용’이 있다고 가정하고,  $t$  시점에서의 크기를  $n$ , 이론  $T_1, T_2$ 에 의해 영향을 받은 이후  $t'$  시점에서의 크기를 각각  $n_1$ 과  $n_2$ 라고 하자. 만약  $n_1$ 과  $n_2$ 가 다르다면, 이는 어떤 이론을 믿느냐에 따라 의식에 현전하는 감각 내용(예. 시각장)의 크기 혹은 정보량 자체가 달라진다는 것이다; 해당 귀결의 참거짓을 따지는 것은 궁극적으로는 경험과학으로 유예할 문제겠지만, 인지적 역량(및 그 역량이 발휘되는 정도)이 고정된 상황에서 받아들이는 과학적 믿음이 달라지는 것만으로 의식되는 정보량 자체가 달라진다는 견해를 받아들이긴 쉽지 않다(시각을 예로 들 경

이러한 점에서 ‘시간 복잡도’를 ad hoc 가설의 기준으로 삼는 견해는 종전의 이론들이 만족시키지 못했던 제약조건 ( $\neg$ ), ( $\neg$ )을 함께 만족시킨다는 장점을 지닌다.

## 4.2. ‘이론의 시간 복잡도’에 대한 심리적 정당화

4.1절에선 과학이론이 어떤 알고리즘/튜링기계와 상응하며, 그에 입각하여 계산할 수 있는 시간 복잡도를 곧 과학이론의 복잡도(단순성)로 간주할 수 있다고 주장하였다. 또한 ‘이론 복잡도’를 ad hoc 가설의 척도로 간주하는 기초를 계승하여 본고에서는 이론/보조가설의 시간복잡도에 입각한 ad hoc 가설의 형식화를 제시하였다.

예컨대 이론 T가 현상 O에 합치하지 않으며, 보조가설 A를 취한 이론 (T+A)와 T를 아예 포기한 대안 이론 T'가 주어지며 이들 모두 O와 합치함이 밝혀졌다고 가정하자. 단, 이때 (T+A)의 시간 복잡도와 T'의 시간 복잡도를 비교하였을 때 후자의 값이 더 작은 경우, 본고의 요지는 곧 A의 추가가 ad hoc한 시도였다고 간주하는 것이다.

---

우 주의attention를 어디에 기울이냐에 초점의 범위는 자그마한 물체에 집중될 수도 있고 시야 전체로 확대될 수도 있지만, 이것만으로 정보량 자체가 달라진다고 볼 근거는 없다). 그러므로 위 귀결을 부정하면 귀류법에 따라  $n_1 = n_2 = f(n)$ 인 함수  $f$ 가 성립하므로, ‘감각내용’의 크기는 이론과 독립적으로 일정하다는 결론이 따라나온다.

- (2) 의미론적 이론적재성: 두 관찰자가 동일한 현상으로부터 각각 다른 관찰명제  $O_1$ 과  $O_2$ 를 얻는다고 가정하자; 이때 두 관찰명제의 관찰용어는 상호번역이 불가능하다. 우린  $O_1$ 과  $O_2$ 를 각각 동일한 형식언어(예. 1차 술어논리의 언어) 구문론에 부합하는 문장  $O_1'$ 과  $O_2'$ 로 번역하는 것을 시도해 볼 수 있다; 이때 관찰용어에 관한 상호번역의 불가능성을 전제하므로, 두 명제가 지니는 비-논리상향의 의미를 비교하는 것은 여전히 불가능하다. 그럼에도 불구하고  $O_1'$ 과  $O_2'$ 는 이제 공통된 구문론을 따르므로, 동일한 기준 하에서 자료의 크기(예. 문자열의 길이, 술어 혹은 원소명제의 개수 등)를 측정해 볼 여지가 생긴다. 이런 방식에 의거하여  $O_1'$ 과  $O_2'$ 의 크기를 측정한 값이 각각  $n_1$ 과  $n_2$ 라고 하자.  $n_1$ 과  $n_2$ 는 실수값이므로,  $n_1:n_2$  비율에 의거하여 시간 복잡도간 비교를 시행시 핸디캡을 반영할 수 있을 것이다.



본고에선 해당 형식화가 제약조건 (⌈), (⌋)을 만족시킨다는 점에서 기존 견해들에 비하여 차별성을 지닌다는 점을 보였지만, 이때 시간 복잡도가 실제 과학적 활동 가운데 무엇을 ‘뜻하는지’를 아직 설명하지 못하였다. 이는 3.2.1절 마지막 단락에서 ‘ad hoc 가설’의 기준이 “이론으로부터의 이반을 야기할만큼 연구-관찰자에게 강력한 동인을 제공”하는 점을 설명해야한다고 요구하는 것과 같은 맥락에 있다.

본고에서는 시간 복잡도가 곧 과학 이론에 입각하여 관찰 현상과의 합치를 보이는 과정, 즉 쿤의 용어를 빌자면 퍼즐풀이 과정에 소요되는 심리적 부담에 상응한다고 주장한다. 이때 심리적 부담이란 ‘연구자가 문제를 풀 때 느끼는 난이도’, ‘문제 풀이에 소요되는 시간’ 등으로 나타난다. 즉 특정한 이론의 시간 복잡도가 현저히 높을 경우, 연구자는 그 이론에 입각하여 문제 풀이를 수행할 때 큰 심리적 부담을 안게 된다; 이러한 심리적 부담 정도가 곧 이론 간의 경합 과정에서 하나의 주요한 변수로 작용한다는 것이다.

이러한 주장을 뒷받침하기 위하여 본고에서는 몇 개의 인지과학적 논의들을 제시하고자 한다. 이러한 기획은 “쿤의 요지는 과학적 이론의 선택에 대한 이해와 평가를 도모하기 위해선 과학적 결과물만이 아니라 과학적 과정에도 주목해야한다는 것 [...] 인간의 이론 평가는 인간 심리에 의존하며, 인간 심리학에 대한 고려없이 이러한 과정에 관해 충분한 설명을 제공할 수 없다”(Brown, 2005, 160)는 과학철학 해석과 입장을 같이 한다.

먼저 본고의 접근 방식처럼 명시적으로 계산 복잡도 이론을 취하는 예시로는 인지 취급용이성 명제 *tractable cognition thesis*에 관한 인지 과학자 이리스 판 로이(Van Rooij, 2008)의 논의를 들 수 있다: 어떤 인지적 역량 *cognitive capacity*이 함수  $\mathcal{W}: I \rightarrow O$ 로 정의된다고 하자; 어떤 알고리즘 수준의 이론  $\mathcal{W}_I: I_I \rightarrow O_I$ 가 있으며 ( $I = I_I$ ) & ( $O = O_I$ ) 일 경우,  $\mathcal{W}_I$ 는  $\mathcal{W}$ 의 진리적 기술 *veridical description*이며,  $\mathcal{W}_I = \mathcal{W}$ 에 대한 여러 층위에서의 설명이 가능해진다(941).

인지 취급용이성 명제는 이렇듯 인지적 역량에 상응하는 이론  $\mathcal{M}$  일체가 취급용이하다(*tractable*)는 주장이다. 취급용이성에 관한 일반적인 정의는 특정 알고리즘의 시간 복잡도가 (결정론적 튜링기계에서) 다항식 시간(*polynomial time*)에 해당하는 경우(예.  $O(n^2)$ ), 오직 그 경우에 해당 알고리즘이 취급용이하다고 보는 것이다(‘Cool-Karp thesis’; 인지과학적 맥락에선 ‘*P-cognition thesis*’). 즉, 인지 취급용이성 명제의 요지는 인간의 인지적 역량 일체라면 결정론적 튜링기계에서 다항식 시간 내에 처리되는 문제보다 더 복잡할 수는 없다는 것이다(946-948; Linz, 2001, 354).

하지만 이러한 결론은 받아들이기 힘들다: 범주화(*categorization*), 유사도(*similarity*) 측정 등 매우 혼란 인간의 인지적 역량들조차 다항식 시간을 초과하는 시간 복잡도를 요구하는 문제들로 알려져 있기 때문이다(Van Rooij, 2008, 954). 이러한 기존의 견해에 맞서 판 로이는 매개변수화 복잡도(*parameterized complexity*)에 기초한 대안적 기준을 제시함으로써, 개선된 형태의 인지 취급용이성 명제를 제시하는데 목적을 둔다(957).

과학적 활동은 인간의 인지적 역량을 통해 이루어지는 활동이며, 본고에선 시간 복잡도가 그런 인지적 과정으로서 과학적 활동이 부담하는 심리적 복잡도에 해당한다고 주장한다. 이때 인지심리학적 관점에서 이루어지는 상기 논의는 시간 복잡도에 입각하여 인지적 역량을 분석한다는 점에서 본고의 논의를 뒷받침한다. 결정적으로 인지 취급용이성 명제는 인간이 지니는 인지적 역량의 최대 한계를 시간 복잡도의 측면에서 규정하고 있다; 이는 지나치게 높은 이론의 시간 복잡도가 곧 인지적 부담을 초래하며, 이론을 포기하는 이유를 제공한다는 본고의 논지와 일치한다.

다만 이때 주의할 점은 인지 취급용이성 명제에서 시간 복잡도를 다항식 시간으로 한정하는 규정을 이론의 시간 복잡도에도 안일하게 소급 적용해서는 안된다는 점이다. 과학 이론과 현상과의 합치를 보이는

문제의 시간 복잡도는 다항식 시간을 초과할 가능성이 높기 때문이다.

물론 과학 이론을 구체적으로 어떻게 모형화시키느냐에 따라 시간 복잡도는 차이가 날 가능성이 있다. 하지만 앞에서 제시한 대표적 모형들을 검토해 보자. 베이즈주의적 관점에서 사후확률을 최대로 하는 이론을 고르는 문제는 다항식 시간 이상을 소요하는 것으로 알려져 있다(954). 아울러 가설-이론 모형을 취한다면 그 뼈대로 삼을만한 논리 체계가 필요한데, 형식 논리체계 중 가장 간단한 명제논리 체계조차 명제의 충족가능성을 보이기 위해선 NP-완비complete 시간 복잡도가 요구된다(Cormen et al., 2009, 1071-7). 상기한바 기준을 취할 경우, 과학 이론과 현상과의 합치를 보이는 문제의 시간 복잡도는 이렇듯 다항식 시간을 초과할 가능성이 크다.

하지만 4.1절에서 밝혔듯이 과학 이론과 관찰 현상 간의 합치를 보이는 문제는 경험적으로 빠르게 해결되는 문제로 보이며 다항식 시간을 뛰어넘는 시간이 걸리지는 않는 것으로 보인다. 이는 판 로이의 논지처럼 대안적 계산 복잡도 이론을 취하는 방식으로든 설명할 수 있겠지만, 다른 한편으로는 과학 이론과 현상간의 합치를 보이는 과제가 매번 완전 탐색exhaustive search으로 수행되는 건 아니라는 점을 지적할 수 있다.

완전 탐색은 간단히 말해서 결정론적 튜링기계가 취하는 문제 풀이 방식으로 모든 경우의 수를 다 따지는 방식이라고 할 수 있으며, 판 로이의 인지적 역량 모델은 이러한 완전 탐색 방식을 전제한다. 하지만 실제 인간의 인지 과정이 완전 탐색을 실시하는지도 의심스러울 뿐더러, 관찰-이론 합치 문제를 푸는 과정은 매번 완전 탐색으로 이루어지는 것은 아니다. 과학 공동체에서는 현시점까지 받아들여진 기존의 이론적 성과들을 전제로 받아들이며, 이를 기반으로 관찰-이론 합치 문제를 푸는 것을 인정하기 때문이다. 예컨대 입자물리학자라고 한들, 농구공이 골대에 튕겨 나오는 과정을 바라볼 때마다 표준모형으로부터 시작하여 그 현상이 일어나기까지의 도출 과정을 일일이 따져볼 필요는 없다. 탐색 문제를 설명할 때 흔히 사용되는 그래프 이론에서의 나무구조tree structure를 본으로 삼아 생각해볼 때, 완전 탐색이 뿌리root로

부터 모든 분기를 다 따지는 것이라면, 실제 관찰-이론 합치 문제는 이미 탐색이 완료된 최말단 노드 *terminal node*부터 탐색을 하는 것을 허용하는 것에 상응한다.

그리고 무엇보다 본고에서 이론의 시간복잡도를 논할 때 취하는 것은 결정론적 튜링기계가 아닌, 비결정론적 튜링기계를 사용한 시간 복잡도이다; 전자가 관찰-이론 합치 여부를 보이는데 사용되는 시간을 소요한다면, 후자는 합치 여부가 보여진 것을 전제할 때 관찰 현상을 설명하기 위해선 몇 차례의 이론적 도출 과정이 요구되는지를 보이는 척도이다. 그러므로 인지 취급용이성 명제의 주장이 발견의 맥락에 관한 것이라면, 본고의 논지는 정당화의 맥락에 관한 것이므로 양자의 주장은 독립적이다.

즉 결론적으로 인지 취급용이성 명제는 이론의 시간 복잡도에 전적으로 적용될 수는 없지만, 적어도 계산 복잡도에 의거하여 심리적 부담을 포착할 수 있다는 지점에서는 본고의 논의와 맥락을 같이한다.

이렇듯 인지 취급용이성 명제가 ‘시간 복잡도’라는 계산 복잡도 이론의 개념을 사용하는 것을 인지심리학적 측면에서 원론적으로 정당화해 준다면, 이를 보다 직관적으로 보여주는 연구의 예시로는 심리언어학에서의 **도출 복잡도 이론** *derivational theory of complexity*(이하 “DTC”)을 들 수 있다.

DTC는 고전적인 변형-생성 문법의 입장에서 문법의 심리적 실체성을 주장하기 위해 고안된 이론이다: 변형-생성 문법에선 자연어 문장의 표층 구조 기저의 심층 구조를 전제하며, 심층 구조 문장의 변형을 통하여 실제 자연어 문장이 도출된다고 주장한다. DTC는 이때 소요되는 변형 절차의 횟수가 곧 해당 문장의 심리적 복잡도의 지표라는 것을 그 요지로 삼는다; 즉 심리언어학적으로 DTC는 특정 문장이 도출되기 위해 요구되는 단계의 수가 실제 문장을 처리/이해하는데 소요되는 시간과 상관성을 지닌다고 주장하였다. 변형-생성문법에 입각한 DTC의 예측은 차후의 여러 실험들에 의하여 반박되었지만, 근래의 보완된 연구는 DTC의 기획이 전적으로 틀리지는 않았음을 시사한다(Carroll, 2008,

38-9). 예컨대 최소주의 이론 등 생성문법 전통의 최신 이론들은 위와 같은 반박에 전적으로 노출된다고는 볼 수 없으며, (약한) 의미론적 합성성이 재귀적 방식으로 설명되는 점을 고려할 때 광의의 DTC는 여전히 인지과학 및 인지신경과학에서 또한 인정할 수밖에 없다는 반대 의견 또한 제기된다(Marantz, 2005, 439-40).

DTC에 얽힌 쟁점은 크게 (1) ‘문장의 도출에 소요되는 단계’를 규정하는 개별 문법 이론이 옳은 것인지 (2) 문장의 도출 과정에 개입하는 기제가 통사론적 기제 하나로 국한되는지 아니면 여럿인지에 관한 문제이다(cf. Sag and Wasow, 2011).

이러한 지적들은 ‘문장의 도출에 소요되는 절차 적용 횟수가 그 복잡도에 대응한다’는 DTC의 전제와 근본적으로 상치되지 않는다: 먼저 (1)의 경우 ‘자연어 문장이 일련의 절차에 걸쳐 도출된다’는 전제를 취하는 한 어떤 문법 이론이든 DTC의 전제와 충돌할 이유가 없으며, 이는 (2) 역시 마찬가지이다; 다만 (2)의 경우 복수의 기제에 의한 도출을 인정한다는 점에서 병행parallel 처리에 대한 고려가 필요한데, 계산복잡도의 측면에서 병행 처리는 다항식 시간 배율 이상의 속도 변화를 가져오지 않는다(Van Rooji, 2008, 967-8).

이러한 상기 논의는 DTC에서 전제하는 바 심리적 복잡도에 관한 전제가 본고의 기획과 맥락을 같이 한다는 점을 보여준다: 양자 모두 특정한 객체(관찰현상/문장)가 체계(과학 이론/문법)에 의하여 도출되기 위해 소요되는 절차의 적용 횟수/단계를 심리적 복잡도에 대한 지표로 삼기 때문이다.

이렇듯 상기 논거들은 본고에서 제시하는바 ‘이론의 시간 복잡도’가 심리적 측면에서도 과학적 활동의 복잡도의 지표가 될 수 있음을 지지한다; 과학적 활동 또한 명백한 인지적 활동이므로, 과학적 활동을 수행할 때의 복잡도는 곧 심리적 부담으로 전이되기 때문이다.

#### 4.3. 형식화 예시: 연역-법칙 모형에서의 시간 복잡도

본고에선 지금까지 과학 이론에 상응하는 튜링 기계의 시간 복잡도를 준거로 삼아 이론의 복잡도를 측정할 수 있고, 이를 토대로 한 ad hoc 가설의 형식화 방안을 제시하였다. 아울러 본고에선 이런 시간 복잡도가 곧 인지적 복잡도에 해당한다고 주장하며, 이를 근거할만한 인지과학적 논거들을 제시하였다.

본고의 논의는 이렇듯 과학 이론에 상응하는 튜링 기계를 준거로 하여 이루어진다; 하지만 과학 이론에 상응하는 튜링 기계를 고안하는 것이 용이한 과제일 것으로 예상되지는 않는다. 그 고안이 이루어질 수 있다는 가능성은 보장될지언정, 고안을 하기 위해서는 이론 전체에 대한 전폭적인 재구성이 요구되기 때문이다.

하지만 본고의 논의는 튜링 기계 그 자체에 한정되는 것이 아닌, 튜링 기계에 상응하는 별도의 체계에도 적용될 수 있을 것으로 보인다; 예컨대 형식 논리체계를 본 뜬 연역-법칙 모형이 그 대표적인 예시다. 비록 튜링 기계에서의 정의만큼 정밀하지는 않을지라도, 우린 이러한 대안적 체계에 입각한 시간 복잡도를 측정할 수 있다. 디지털 컴퓨터 프로그램의 시간 복잡도를 계산할 때, 꼭 그 프로그램에 해당하는 튜링 기계를 참조하는 것이 아니라 추상화된 고급 언어의 코드를 분석의 대상으로 삼는 것은 이와 같은 맥락에 있다.

즉 본고의 논의는 특정한 과학 이론 모형에만 배타적으로 적용되는 것이 아니라, 기계적/효과적 계산가능성을 인정하는 과학 이론 모형 일체에 적용될 수 있다. 본고에서는 이러한 전제 하에서 연역-법칙 모형을 예시로 들어 본고의 논의가 어떻게 적용되는지를 설명하고자 한다.

연역-법칙 *deductive-nomological* 모형은 과학적 설명의 대표적인 모형 중 하나로, 전제 조건 및 이론 법칙들을 기술하는 설명항과 관찰 형상을 기술하는 피설명항간에 논리적 귀결 관계가 성립하는 것을 요구하는 입장에 해당한다(Hempel and Oppenheim, 1948; Hempel, 1966,

49-53). 특히 이는 가설-이론 모형과 함께 쓰임으로써, 과학 이론이 서로 논리적 도출 관계를 이루는 명제들로 구성된 연역 체계처럼 작동하게끔 한다. 이러한 연역-법칙 모형은 본래 의도된바 과학적 설명의 측면에서 인과성을 충분히 설명하지 못한다는 등의 비판이 제기되나 (Bromberger, 1966), 적어도 본고에서 주목하는바 ‘관찰 현상과 이론 간의 합치’를 설명하는 데는 별 문제가 없어 보인다.

이렇듯 과학 이론이 연역 체계의 형식으로 재구성될 수 있다고 전제할 경우, ad hoc 가설이 이론에 임의적으로 부가되는 것을 허용하는 것은 해당 이론이 흡사 건전한 이론 체계처럼 기능하는 것을 가능하게 한다: 건전한 논리 체계는 그로부터 증명가능한 모든 명제들이 곧 타당한 명제들인 체계로 정의된다(Sider, 2010, 63). 만약 과학 이론에 ad hoc 가설이 임의적으로 추가될 수 있을 경우, 특정 이론 T는 관찰현상 A에 불합치 할 경우에도 ad hoc 가설을 통하여 A와도 합치하는 T'로 임의적인 변화가 가능하다; 이러한 과정이 계속될 경우, 이론 T는 ad hoc적 수정만을 통하여 건전한 체계인양 작동하게 되는 것이다. 이렇듯 1절에서 소개한 바 ad hoc 가설의 문제점은 연역-법칙 모형에서 명백하게 드러난다.

그렇다면 이러한 연역-법칙 모형에 관하여 본고에서 제시한 ‘이론의 시간 복잡도’는 어떻게 적용될 수 있는가? 이는 ‘주어진 이론 체계 T에서 주어진 A를 도출하기까지 소요되는 도출 단계의 횟수’에 해당한다고 볼 수 있다; 즉 주어진 현상 A에 관하여 유사-건전한 이론 T1과 이론 T2가 경합중이라고 할 때, 동일한 조건이 주어졌다는 전제하에서 ‘T1 → A’를 보이기까지 총  $m$ 단계, ‘T2 → A’를 보이기까지 총  $n$ 단계가 소요되며  $m < n$ 인 경우, A에 관하여 이론 T1이 더 단순하다고 주장하는 것이다.

이는 여러 차례 소개한 프톨레마이오스 천문학과 코페르니쿠스/케플러 천문학간의 경합을 예화로 들 때 쉽게 이해될 수 있다. 상기했다시피 두 천문학 이론은 당대의 관찰 결과들을 도출시키는데 성공할 수 있었다; 다만 전자가 이론이 개진되면 개진될 수록 주전원 가설을 거듭

도입하여 고안해 낸 복잡한 궤도식을 바탕으로 관찰 결과를 이론에 합치시키는데 주력하였다고 하면, 후자는 오히려 지동설 및 타원 궤도의 법칙 같은 보다 강력한 법칙에 의거하여 적은 수의 도출 단계만으로도 관찰 결과와의 충돌을 방지할 수 있는 것이다. 이러한 설명은 3.2.1에서 소개한바 프톨레마이오스 천문학자들이 불만을 느꼈던 프톨레마이오스 천문학의 복잡도에 대한 진술과 부합하는 것으로 보인다.

또한 1절에서 언급한 ‘차고 속의 용’ 예화에 대해서도 동일한 방식의 설명이 가능하다:  $T1 = \text{‘차고 안에 용이 없다’}$ ,  $T2 = \text{‘차고 안에 용이 있다’}$ ,  $B = (\text{일상적인 인식론적 전제들})$ 이라고 정의하자. ‘ $T2 \& B$ ’로부터  $E1 = \text{‘용이 보이지 않는다’}$ 를 도출시키기 위해선  $A1 = \text{‘용이 투명하다’}$ 라는 보조가설이 도출 과정에서 추가적으로 개입해야 한다. 아울러 관찰이 더 이루어짐에 따라  $E2 = \text{‘용은 보이지도 않고 온도 탐지기에도 안 나타난다’}$  등을 ‘ $T2 \& B$ ’로부터 도출시키기 위해선  $A1$ 에 더불어 또 다른 보조가설  $A2 = \text{‘용은 주변 온도와 완전히 동화되어 있다’}$ 라는 새로운 보조가설이 동원되어야 한다. 즉, 보조가설  $A1, A2 \dots$  이 추가로 도입됨에 따라 이론의 시간 복잡도는 점점 늘어나며, 이는 이론에 대한 심리적 부담의 증가에 상응한다. 반면 ‘ $T1 \& B$ ’로부터  $E1, E2 \dots$  을 도출시키기 위한 시간 복잡도는 고정적이다. 따라서 ‘ $T1 \& B$ ’와 ‘ $T2 \& B \& A1 \dots$ ’이 결합할 때, 후자를 ad hoc하다고 분류하는 것이 가능하다.

이러한 방식은 논리적 도출 횟수에 입각한 심리적 복잡도 측정 기준을 제시하는 DTC의 기획과 정확히 일치한다. 나아가 이러한 문제 해결 방식은 상용 컴퓨터 언어에서도 찾아볼 수 있다: 컴퓨터 언어 Python은 함수의 재귀적 호출을 허용하는 고급 언어이다. 하지만 Python 인터프리터는 프로그램 실행 과정 중에서 아무런 문제가 나타나지 않더라도, 해당 프로그램에서 재귀 호출 횟수가 일정 숫자를 넘을 경우 프로그램 실행 중 오류가 발생한 것으로 판정한다. 논리/산술적 측면에서 재귀 호출의 숫자를 일정하게 제한할만한 근거는 없다; 그럼에도 Python 인터프리터가 이러한 장치를 도입하는 까닭은 재귀 횟수가 일정한 횟수를 넘어갈 경우, 이는 정상적인 프로그램이 아니므로 스택 오버플로우를



야기할 개연성이 크다고 판단하기 때문이다<sup>19)</sup>. 이런 여러 발상들의 공통점은 논리적 타당성이 보장된다 하더라도, 논리적인 도출 과정의 연쇄 길이 자체를 이론/프로그램의 오류 여부에 관한 판단의 근거로 삼는다는 점이다.

물론 상기 방법론은 여러 불분명한 문제점들을 지닌다; 과학 이론과 상응하는 튜링 기계가 있을 때 있을 때 그와 완벽하게 상응하는 연역-이론 모형을 만들 수 있는 가능성은 보장될지라도, 실제 어떻게 만들 수 있는지 방법에 관해서는 아직 밝혀진 바가 적기 때문이다.<sup>20)</sup>

먼저 연역-이론 모형의 뼈대가 되는 논리 체계와 도출 방식을 어느 것으로 삼는지가 문제가 된다: 예컨대 1차 술어 논리 체계는 결정 불가능한 체계이므로 채택할 수 없으며, 결정가능한 논리 체계를 선택한다 할지라도 어떠한 도출 방식을 취하는지(예. 자연 연역, 공리적 증명방식 등)에 따라 도출 단계의 수가 크게 차이가 날 수도 있기 때문이다. 따라서 이론/보조가설간의 복잡도를 비교하기 위해선 서로 경합하는 이론들 간에 각종 조건들이 동일하게 유지되어야만 한다.

아울러 상기 설명에선 도출 과정에서 사용되는 모든 법칙들이 동등한 것으로 간주된 반면, 실제 이론을 평가할 때에는 각 법칙들마다 도출 과정에서의 가중치를 다르게 매겨야 할 것으로 보인다. 예컨대 명제논리에서의 Cut Rule을 사용할 경우, 복수의 삼단논법 사용을 요구하는 풀이과정의 하나의 도출 단계로 압축될 수 있기 때문이다(Sider, 2010, 73). 이렇듯 각 법칙마다 다르게 매겨지는 가중치는 그 법칙에 대응하는바 튜링 기계 상의 이동 횟수와 상관성을 지닌다.

그럼에도 “합리적인 [계산] 기계들은 다항식 시간 및 상수 인자공간

19) Python 3.3.2 (v3.3.2:d047928ae3f6, May 16 2013, 00:06:53):

<http://docs.python.org/2/library/sys.html>

20) 정의상 연역-법칙 모형은 확률적 법칙이 포함되는 이론을 수용하지 못하므로, 그 한계는 본 방법론에도 동일하게 적용된다. 차후 연구가 필요하겠지만, 이러한 한계점은 험펠이 연역-법칙 모형과 함께 제시한 귀납-통계(inductive-statistical) 모형 등을 동원하면 충분히 극복할 수 있지 않을까 추측된다.

부하overhead 안에 서로를 재현emulate할 수 있다”는 불변성 명제 invariance thesis를 받아들일 경우(Boas, 1990), 튜링 기계에 상응하는 여러 형식 체계들의 시간 복잡도간의 차이는 점근적 관점에서 현격하진 않다. 따라서 이론의 시간 복잡도가 그 구체적인 형식화 방식에 따라 전적으로 좌지우지될 것이라는 비관적인 의견을 품을 필요는 없는 것으로 보인다.

이런 여러 잠재적인 해결사항들을 고려할 때, 연역-법칙 모형은 아직 그 구체적인 구현 방식에 관해서는 더 많은 연구가 필요할 것으로 보인다. 하지만 이런 시도가 전통적인 논리 경험주의의 모형이며, 전적으로 새로운 시도가 아닌 것임을 고려할 때, 상기 방식에 입각한 ad hoc 가설 해명은 충분히 가능한 시도일 것으로 보인다.

## 5. 나가며

이상 본고에서는 관찰의 이론적재성을 전제한 상태에서 ‘ad hoc 가설’ 개념이 어떻게 형식화될 수 있는지, 그 해당 여부를 어떻게 가릴 수 있는지에 관한 논의를 전개하였다. 본고에서는 험펠, 포퍼 및 쿤 등의 고전적 이론들과 더불어 통계적 단순성에 입각한 견해, 베이즈주의 등 근래에 제기된 ad hoc 가설에 관한 견해들을 함께 검토하였다. 하지만 본고에서는 이들 견해들이 기존의 이론내재성 및 관찰의 이론적재성이라는 두 가지 조건을 만족시키지 못함을 보였다. 이후 본고에서는 해당 조건들을 만족시키는 대안적 기준으로 ‘이론의 시간 복잡도’를 제시하였으며, 이러한 기준을 뒷받침하는 인지과학적 근거들 및 연역-이론 모형으로서의 응용 예시를 설명하였다.

본고의 첫 번째 의의는 상술하였다시피 ‘ad hoc 가설’에 관한 대안적 기준을 제시하였다는 점이다: ad hoc 가설에 관한 기존의 대부분의 기준들은 그 정도의 차이는 있을지언정 객관적 관찰 자료와의 부합 정도

를 참조하였다. 이는 쉽게 받아들임직한 견해임에도 불구하고, 쿤의 이론 등 관찰의 이론적재성을 전제하는 유력한 과학철학적 견해와는 부합하지 않는다. 반면, 쿤 등 이론적재성을 받아들이는 기존의 견해들은 ad hoc 가설에 관한 체계적인 설명을 시도하지 않은 것으로 보인다.

본고의 견해는 2절에서 논한바 험펠의 두 기조 중 ‘이론의 복잡도’를 계승하는 입장이다. 본 입장은 ‘ad hoc 가설’에 대한 체계적인 설명을 제공함과 동시에 4.2절에서 논한 바처럼 관찰의 이론적재성과 충돌하지 않는다는 장점을 지닌다. 또한 이는 ad hoc 가설의 형식화에만 국한되지 않고, ad hoc 가설의 도입이 이루어짐에 따라 기존 이론의 교체가 이루어지게 되는 심리적 동기를 설명할 수 있다.

이러한 측면에서 강조될 점은 본 논의가 이론적재성을 전제할 때 생겨나는 문제를 해결하는데 유효하지만, 결코 이론적재성을 전제하는 상황에만 배타적으로 유효한 것은 아니라는 점이다: 본고의 기획은 이론적재성으로 인해 비합리주의적 귀결이 도출됨직한 ‘최악의’ 경우조차도 여전히 합리적 설명이 가능하다는 것을 보여주고자 하는데 있는 것이지, 본고의 논의가 그런 최악의 상황에만 유효하다는 것을 의미하지 않는다. 더욱이 본고에선 기존의 시도들이 이론의 관찰적재성을 고려하지 않은 점은 제외하더라도(조건(ㄱ)), 여전히 다른 문제들을 노출시키는 점을 지적하였다. 본고의 논의는 이러한 기존의 여러 문제들로부터 자유롭다. 그러므로 이론적재성을 받아들이지 않을 때에도 본고의 논의는 기존의 여러 문제들을 해결하는데 유효하다. 따라서 본고의 논의는 이론적재성을 거부할 때에 고안될 법한 이론들과도 결코 배치되는 것이 아니라, 오히려 상보적이다.

두 번째로 본고의 논의는 이론의 우아함(*elegance*) 혹은 설명적 단순성에 관한 새로운 기준을 제시한다: 우아함(설명적 단순성)이란 이론이 자연을 설명하는 방식의 복잡/단순성을 일컫는 용어로, 지금까지 본고에서 언급한 “복잡도(단순성)”는 모두 우아함에 해당한다(Fitzpatrick; Baker, 2013). 기존의 이론적 우아함에 관한 논의들이 매개변수의 개수, 이론에서 상정된 공준/가정들의 수처럼 이론의 정적인 측면에 주목하였다면, 본고에서는 그런 공준/가정들이 관찰 현상들을 설명할 때 실제

‘활용되는 횟수’를 기준으로 한다는 점에서 기존 이론들과 차별성을 지닌다.

마지막으로 본고의 기획은 과학철학의 유력한 전통중 하나인 비합리주의에 대한 적극적인 반대 연구로서의 의의를 지닌다: 관찰의 이론적재성은 관찰 현상의 객관성을 부정하게끔 한다는 점에서 상대주의적 귀결을 지니므로, 과학의 합리성에 대한 의구심을 낳는다; 이는 관찰의 이론적재성을 주장한 쿤이 비합리주의라는 비판을 받은 이유이기도 하다. 본고는 관찰의 이론적재성을 전제함에도 불구하고 이론내재적인 ad hoc 가설 판단 기준이 세워질 수 있다는 것을 보임으로써, 이렇듯 매우 상대주의적인 전제가 가정될 때조차도 여전히 과학에 대한 합리적 설명이 가능하다는 사실을 보여준다는 의의를 지닌다.

그러나 본고의 논의는 ‘ad hoc 가설’에 대한 설명에 있어서 여전히 많은 부족한 점을 지니며, 아직 해결되지 못한 차후 연구 주제들을 남겨두고 있다.

첫 번째로는 본고의 논의가 **이론 수정 자체에서 야기되는 시간 복잡도**를 충분히 설명하지 못했다는 점을 들 수 있다: 상술한 바처럼 본고의 요지는 과학자가 이론을 접할 때 느끼는 인지적 부담이 이론의 시간 복잡도에 대응한다는 것이다. 이때 본고에선 이론 T에 보조가설 A가 도입 되어 (T+A)가 될 경우, (T+A)의 시간 복잡도에 주목하였다. 하지만 실제 과학 활동의 경우 과학자의 심리적 부담은 갱신이 완료되고 난 결과로서 이론이 나타내는 복잡도 뿐만 아니라, 이론의 갱신 과정 자체에 의해서도 증대될 수 있다; 즉 (T+A)의 시간 복잡도만이 아니라 ‘+A’ 과정 자체의 시간 복잡도 또한 고려해야만 한다는 것이다.

기계학습 분야에서 과도적합 문제에 대처하는 방식은 이러한 문제 의식을 잘 반영한다: 감독 기계학습 알고리즘은 일반적으로 예측상의 오류를 교정하는 방식으로 자료에 대한 학습을 시도한다. 하지만 오류에 지나치게 민감하게 반응하는 것은 상기한 바 과도적합 문제를 낳으며, 이를 방지하는 가장 쉬운 해법중 하나는 오류에 규칙화(regularization) 항을 더하는 것이다. 이때 규칙화 항의 역할은 기존의 모형이 오류에 의하여

교정되는 정도가 지나치게 클 경우, 그 정도를 약화시키는 것이다; 오류에 반응하여 모형이 변동하는 정도가 지나치게 클 경우, 그 복잡도가 증가함에 따라 상기한 과도적합 문제가 발생하기 때문이다(Bishop and Nasrabadi, 2006, 256).

기계학습에서의 ‘오류에 의한 교정’에 정확히 대응하는 것이 바로 상기 도식의 ‘+A’ 과정이다. 즉 기계학습에서 ‘오류에 반응하여 모형이 변동하는 정도’를 억제하고자 하는 것은 ‘+A’ 과정에서 야기되는 시간 복잡도를 최소화하고자 하는 것과 같다.

이러한 문제점은 본고에서 취하는 엄밀한 ‘시간 복잡도’를 통해서 설명할 수 있다. 모형의 변동 정도가 크다는 것은 기존 매개변수들의 값에 절대값이 큰 수가 더해진다는 것이며, 튜링 기계에서 덧셈을 수행할 때 이동 횟수는 그 덧셈항의 절대값 크기와 비례하기 때문이다(Linz, 2001, 233-234).

이렇듯 이론 수정 자체에서 야기되는 시간 복잡도의 증가는 본고의 논지와 합치하는 방식으로 설명될 수 있지만, 이 문제가 어떻게 보다 좋은 모형을 통하여 제시될 수 있을지에 관해서는 더 많은 연구가 필요할 것으로 보인다. 본고에서의 ‘ad hoc 가설’ 판단은 기존 이론 T와 관찰 현상 E으로부터 도출될 수 있는 복수의 이론 후보간의 경합을 전제하는 반면, 많은 기계학습 알고리즘들은 결정론적이므로 (T, E)로부터 하나의 이론만이 도출되는 것을 전제하기 때문이다.

연역-이론 모형처럼 본 문제에 관해서도 적절한 모형을 찾을 수 있을 경우, 이는 본고의 논지와 상호보완적으로 이론 복잡도 및 ad hoc 문제에 관하여 보다 완결적인 설명을 제공할 수 있으리라 기대된다.

본고에서 고려하고 있지 못한 두 번째 요소로는 이론의 절약성 *parsimony* 혹은 존재론적 단순성이 있다: 이는 이론에서 전제하는 존재자 혹은 존재자의 유형의 개수를 기준으로 한 척도이다(Fitzpatrick; Baker, 2013). 절약성은 상기한 우아함과 더불어 이론의 복잡도를 구성하는 주요한 개념이지만, 해당 척도가 본고의 관심대상인 인지적 부담에 영향을 미치는지, 미친다면 어떤 영향을 미치는지에 관해서는 충분한 검토를 하

지 못하였다. 다만 현재로서 추측할 수 있는 것은 이론의 우아함이 시간 복잡도와 밀접한 관계를 지니고 있다면, 이론의 절약성은 계산 복잡도 이론의 또다른 제재인 공간 복잡도 *space complexity*와 연관되어 있음직하다는 점이다.

마지막으로 본고에서 다루지 못한 기존 연구들을 언급하고자 한다:

제렛 레플린(Leplin, 1975)은 총 다섯 개의 기준에 의거하여 ‘ad hoc 가설’ 여부를 판단할 수 있다고 주장하는 절충적 견해를 제시한다: 하지만 해당 기준들은 명백하게 관찰 자료와 이론 간의 부합 정도를 준거로 삼는다는 점에서 본고의 제약조건 (L)에 위배된다. 해당 논문에 관한 보다 자세한 반박은 헌트의 논문(Hunt, 2012)에서 확인할 수 있다.

그리고 이론의 단순성을 형식적으로 측정하고자 하는 근래의 시도들로 콜모고로프 복잡도 *Kolmogorov complexity*에 의거한 최소 기술 길이 *minimum description length* (cf. Rissanen, 1999), 최단 전언 길이 *minimum message length* (cf. Wallace and Dowe, 1999) 등의 척도를 동원하고자 하는 견해들이 있다. 콜모고로프 복잡도는 주어진 부호화된 자료가 있을 경우, 이를 출력할 수 있는 최단 프로그램 코드의 길이를 가리킨다(Li and Vitányi, 2009). 해당 척도들은 어디까지나 관찰 자료에 준거한다는 점에서 제약조건 (L)을 위배할 것으로 추측되나, 추후 연구에서보다 자세한 검토가 이루어져야할 것으로 보인다.

또한 본고에서는 충분히 많은 과학사적 사례를 구체적으로 검토하지 못했다는 단점을 지닌다. ‘ad hoc 가설’에 관한 추후 연구는 보다 다양한 과학사적 사례를 검토하는 것과 더불어, 상기한 미결 문제들에 대한 해결을 통해 진일보 할 수 있을 것으로 보인다.

## 참고문헌

- Baker, Alan 2013. "Simplicity". In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2013 ed.
- Bamford, Greg (1999) "What Is the Problem of Ad Hoc Hypotheses?", *Science & Education*, 8(4), pp. 375-386.
- Bishop, Christopher M and Nasser M Nasrabadi (2006) *Pattern Recognition and Machine Learning*. springer New York.
- Bloor, David (1991) *Knowledge and Social Imagery*. University of Chicago Press.
- Boas, Peter van Emde (1990) "Machine Models and Simulations", *Handbook of Theoretical Computer Science (Vol. A)*: MIT Press, pp. 1-66.
- Brock, William A. and Steven N. Durlauf (1997) *A Formal Model of Theory Choice in Science*: Santa Fe Institute. Available at: <http://econpapers.repec.org/RePEc:wop:safire:97-04-031e>.
- Bromberger, Sylvain (1966) "Questions", *The Journal of Philosophy*, pp. 597-606.
- Brown, Harold I. (2005) "Incommensurability Reconsidered", *Studies in History and Philosophy of Science Part A*, 36(1), pp. 149-169.
- Carroll, David W. (2008) *Psychology of Language*. 5th edn.: Thompson/Wadsworth.
- Chen, Xiaoping (2006) "Bayesian Test and Kuhn's Paradigm", *Frontiers of Philosophy in China*, 1(3), pp. 491-505.
- Copeland, B. Jack 2008. "The Church-Turing Thesis". In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Fall 2008 ed.
- Cormen, Thomas H., Charles E. Leiserson, et al. (2009) *Introduction to Algorithms*. 3rd edn.: The MIT Press.
- Earman, John (1992) *Bayes or Bust? A Critical Examination of Bayesian Confirmation Theory*. Cambridge, MA: MIT Press.

- Elgin, Mehmet (2008) "Theory-Laden Observation and Incommensurability", *ORGANON F*, 15(1), pp. 3-19.
- Farmakis, Lefteris (2008) "Did Tom Kuhn Actually Meet Tom Bayes?", *Erkenntnis*, 68(1), pp. 41-53.
- Feyerabend, Paul K. (1962) "Explanation, Reduction and Empiricism", in Maxwell, H. F. & G (eds.) *Scientific Explanation, Space, and Time*, (Minnesota Studies in the Philosophy of Science, Volume Iii).
- Feyerabend, Paul K. (1993) *Against Method*. 3rd edn.: Verso.
- Fitzpatrick, Simon "Simplicity in the Philosophy of Science". *Internet Encyclopedia of Philosophy*. (Accessed on Jan. 23, 2014)
- Forster, Malcolm and Elliott Sober (1994) "How to Tell When Simpler, More Unified, or Less Ad Hoc Theories Will Provide More Accurate Predictions", *The British Journal for the Philosophy of Science*, 45(1), pp. 1-35.
- Franklin, A., M. Anderson, et al. (1989) "Can a Theory-Laden Observation Test the Theory?", *The British Journal for the Philosophy of Science*, 40(2), pp. 229-231.
- Hawkins, Douglas M. (2003) "The Problem of Overfitting", *Journal of Chemical Information and Computer Sciences*, 44(1), pp. 1-12.
- Hempel, Carl G and Paul Oppenheim (1948) "Studies in the Logic of Explanation", *Philosophy of science*, 15(2), pp. 135.
- Hempel, Carl G. (1966) *Philosophy of Natural Science*. Englewood Cliffs, N.J., Prentice-Hall.
- Howson, Colin and Peter Urbach (2006) *Scientific Reasoning: The Bayesian Approach*. 3rd edn.: Open Court.
- Hoyningen-Huene, Paul (1993) *Reconstructing Scientific Revolutions: Thomas S. Kuhn's Philosophy of Science*. University of Chicago Press.
- Hunt, J. Christopher (2012) "On Ad Hoc Hypotheses", *Philosophy of*



- Science*, 79(1), pp. 1-14.
- Jeffreys, Harold (1998[1961]) *The Theory of Probability*. Oxford University Press.
- Kieseppä, I. A. (2001) “Statistical Model Selection Criteria and Bayesianism”, *Philosophy of Science*, 68(3), pp. S141-S152.
- Kuhn, Thomas S. (1970) “The Structure of Scientific Revolutions, 2nd”, *Chicago: Univ. of Chicago Pr.*
- Lakatos, Imre (1980) *The Methodology of Scientific Research Programmes: Volume 1: Philosophical Papers*. Cambridge University Press.
- Leplin, Jarrett (1975) “The Concept of an Ad Hoc Hypothesis”, *Studies in History and Philosophy of Science Part A*, 5(4), pp. 309-345.
- Li, Ming and Paul MB Vitányi (2009) *An Introduction to Kolmogorov Complexity and Its Applications*. Springer.
- Linz, Peter (2001) *An Introduction to Formal Languages and Automata*. 3rd edn.: Jones & Bartlett Publishers.
- Marantz, Alec (2005) “Generative Linguistics within the Cognitive Neuroscience of Language”, *The Linguistic Review*, 22(2-4), pp. 429-445.
- “Nondeterministic Turing Machine”. *Encyclopedia of Mathematics*. Available at: [http://www.encyclopediaofmath.org/index.php?title=Nondeterministic\\_Turing\\_machine&oldid=30560](http://www.encyclopediaofmath.org/index.php?title=Nondeterministic_Turing_machine&oldid=30560). (Accessed on Apr. 3, 2014)
- Oberheim, Eric and Paul Hoyningen-Huene 2013. “The Incommensurability of Scientific Theories”. In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Spring 2013 ed.
- Popper, Karl Raimund Sir (1959) “Discussions: Testability and ‘Ad-Hocness’ of the Contraction Hypothesis”, *British Journal for the Philosophy of Science*, 10(37), pp. 50-50.
- Popper, Karl Raimund Sir (2002[1959]) *The Logic of Scientific Discovery*. Taylor & Francis.

- Quine, Willard V. O. (1951) "Two Dogmas of Empiricism", *Philosophical Review*, 60(1), pp. 20 - 43.
- Rissanen, J. (1999) "Hypothesis Selection and Testing by the Mdl Principle", *The Computer Journal*, 42(4), pp. 260-269.
- Sag, Ivan A. and Thomas Wasow (2011) "Performance-Compatible Competence Grammar", *Non-transformational syntax: Formal and explicit models of grammar*, pp. 359-377.
- Salmon, Wesley C. (1990) "The Appraisal of Theories: Kuhn Meets Bayes", *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1990, pp. 325-332.
- Schindler, Samuel (2013a) "Theory-Laden Experimentation", *Studies in History and Philosophy of Science Part A*, 44(1), pp. 89-101.
- Schindler, Samuel 2013b. "Observation and Theory-Ladenness". In: Kaldis, B. (ed.) *Encyclopedia for Philosophy and the Social Sciences*. Sage.
- Scott, Sarah Louise (2012) "Moving Too Fast: The Making of an Ad Hoc Hypothesis", *Metatheoria - Revista de Filosofia e Historia de la Ciencia*, 2(1), pp. 37-60.
- Sider, Theodore (2010) *Logic for Philosophy*. Oxford University Press Oxford.
- Strevens, Michael (2001) "The Bayesian Treatment of Auxiliary Hypotheses", *British Journal for the Philosophy of Science*, 52(3), pp. 515-537.
- Thornton, Stephen 2013. "Karl Popper". In: Zalta, E. N. (ed.) *The Stanford Encyclopedia of Philosophy*. Spring 2013 ed.
- Van Rooij, Iris (2008) "The Tractable Cognition Thesis", *Cognitive Science*, 32(6), pp. 939-984.
- Wallace, C. S. and D. L. Dowe (1999) "Minimum Message Length and Kolmogorov Complexity", *The Computer Journal*, 42(4), pp. 270-283.